IMPERIAL COLLEGE

Final Report for Objective 3

Critical Evaluation of Evidence Concerning the Empirical Form of Dose-Response Relationships at Low Dose Exposure to Genotoxic Carcinogens in Animals

FSA Project: T01051
INTERPRETATION OF MARGINS OF EXPOSURE FOR GENOTOXIC CARCINOGENS

30/10/2012

Report by:

Ehi Idahosa-Taylor

Contributors/ project participants:

Alan Boobis, Lesley Rushton, Andy Hart, Villie Flari, Peter Craig, Anjali Mazumder

This report details the systematic review of toxicological evidence regarding the form of the dose-response curve for genotoxic carcinogens at low exposures in animals and the implications for a level of concern for Margins of Exposure.

Table of contents

I.   Introduction

Cancer has historically being considered to be a genetic disorder characterised by the observed mutations in genes that are involved in the regulation of growth [3].

On the basis of studies on the effects of ionizing radiation the *one hit* theory for carcinogenesis was developed, that it could take only a single interaction of an ionizing particle to initiate DNA mutation [4]. By analogy, reactions between DNA and genotoxic chemicals are considered to occur in a "stochastic" manner, which similarly implies that a single nuclear DNA damaging event could result in a carcinogenic response [5]. This led to the view that any exposure to a carcinogen could result in cancer – commonly referred to as the "one hit theory for carcinogenesis". A consequence of this was the presumption that a chemical that was carcinogenic in rodents could pose a risk to humans at any level of exposure. Hence, from a regulatory point of view, the determining factor in managing exposure to carcinogens is the ability of such chemicals to react with DNA.

Historically, the regulation of carcinogenic chemicals rested on the basis of policies passed within the Federal Agencies in the United States (US). Championed by the then Congressman J. Delaney, the Delaney Clause was passed as an amendment to the Federal Food, Drug, and Cosmetic Act in 1958 [6]. The Delaney Clause stated that "*no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal, or if it is found, after tests which are appropriate for the evaluation of the safety of food additives, to induce cancer in man or animal.*" While the intention of the Clause was to protect the American people from the increasing incidences of cancers, because little was known about the causes of cancer at this time [6, 7], this ruling for the prevention of human exposure to potential carcinogens was based on opinion rather than on scientific evidence [8].

On the basis of the foregoing, classification of a chemical as a genotoxic carcinogen has proved to be a source of difficulty for risk managers. In cases where the substance cannot be banned or substituted, an approach most favoured by agencies in the US is the use of linear extrapolation to dose levels that are considered to be virtually safe (the "virtually safe dose"), i.e. associated with a low probability of concern for human safety (used by the USEPA as a level of risk corresponding to an excess of one case of cancer among a million individuals over a lifetime of exposure, defined as 70 years). This approach facilitated a change in risk management practices, particularly by the United States Environmental Protection Agency (USEPA) (at least in relation to pesticide residues in food), which permitted exceptions to the Delaney clause on the basis of considerations of risk/benefit analysis [9], where exposure to carcinogens based on linear extrapolation were at or below a "virtually safe dose".

In contrast, the approach most favoured by agencies in the UK is the reduction of the potential level of exposure of genotoxic carcinogens to as low as reasonably achievable (ALARA). However, this provides no information to enable prioritisation of risks and hence

recently, the margin of exposure approach has been proposed for this purpose and to improve risk communication [10].

1.1 *Dose Response Analysis (model types)*

The proposal for an alternative to the position mandated by the Delaney Clause was originally put forward by Mantel and Bryan [11] who suggested the calculation of an upper confidence limit for observed tumours. The authors proposed the use of modelling methods whereby the dose response curve would be extrapolated from the high doses where responses to the effects of carcinogens were observed in experimental animals towards zero. It was proposed that this would enable the calculation of a "virtually safe dose" (VSD) for a given arbitrary low response (e.g. 10^{-8}) [11, 12]. The authors proposed the use of a probit response line which was considered to be conservative. This method was later revised to enable the consideration of background tumour incidences and extrapolation from experiments using several doses [12, 13].

The risk assessment of human cancer from carcinogenic chemicals is based on animal experimental data at pre-determined dose exposure levels. This is highly limiting due to the inadequacy of the experimental data to provide information at low dose exposure levels, thereby requiring modelled extrapolations many orders of magnitude below observable data, and often with too few sacrificial or premature expiration of animals to assess time to tumour. Various characterisations and model fitting techniques have been employed to describe the dose-response relationship which is an important aspect of risk assessment. The models commonly employed when assessing dose-response relationships are categorised into three classes discussed in more detail below. For the class of model, a brief description of their properties, structure and assumptions are given.

Three classes of models:

i. Linear (no-threshold) Model

The linear models, also referred to as the no-threshold models, are the simplest and most frequently applied for low dose extrapolation. The linear no-threshold model assumes that there is theoretically no level of exposure from a carcinogen that does not pose a small, finite, probability of generating a carcinogenic response. It assumes that the response is directly proportional to the dose at all dose levels, i.e. the carcinogen is considered harmful with no safety threshold and the sum of several very small exposures have the same effect as one larger exposure (additive). The extrapolation phase of this type of assessment uses a straight line from the point of departure for the observed data to the origin, taking the following form:

$$r \propto cd + b,$$

where r is the tumour response, c is the 'cancer' slope, d is the dose level and b is a constant (zero response).

It is easily applied, to extrapolate the expected number of deaths caused by exposure to the carcinogen. A quantity of the carcinogen can be translated into a number of deaths without any adjustment for the distribution of the exposure. If a given dose of carcinogen is found to produce one extra case in every thousandth person exposed, the model projects that one thousandth of this dose will produce one extra case in every million people exposed, i.e. any given quantity of the chemical will produce the same number of cancers no matter how thinly the dose rate is spread.

As risk is generally expressed as the probability that an individual will develop cancer by some age, for a population exposed to different doses, the expected number of cases is

$$E\{N(r)\} = \sum_i N_i \Pr(d_i),$$

where N_i is the number of exposed at dose d_i and $\Pr(d_i)$ is the probability of developing cancer at that dose level [14, 15]. Thus, this approach is used to estimate doses that correspond to specified levels of allowable risk, e.g. risk $< 1/100,000$. The main criticism of this extrapolation approach is that the estimation of the risks for developing cancer is considered to be very conservative.

ii.   Tolerance Distribution Model

Tolerance distribution models used in low dose extrapolation work under the assumption that every individual animal tested within the study population has its own level of tolerance or threshold for the exposure to the chemical being investigated. The variations in the threshold exhibited by individuals within the study population are described in terms of a cumulative probability distribution function. An example of this type of dose response model is the log-probit model proposed by Mantel and Bryan [11, 13]. The log-probit model facilitates the calculation of an upper bound estimate of the dose that is associated with a predefined acceptable level of cancer risk through extrapolation from the observed portion of animals that have exhibited tumours along a dose response line with a slope of one [16].

The tolerance distribution model class opposes the above competing linear no-threshold model class school of thought on the premise that very small exposures are harmless. The tolerance distribution (or threshold) model, taking on a *non-linear response assessment*, assumes that a range of exposure between zero to some finite value can be tolerated with essentially a zero chance of expressing a carcinogenic response; the threshold is the dosage point from which risk of response occurs, and prudently is set for the most sensitive members of populations.

The non-linear term in characterising the dose-response relationship is best described as a sigmoid shape in which the slope is zero below a low-threshold dose (i.e. no response) and above a high-threshold (e.g. lethal dose). This simply characterises the `endpoints' or thresholds of dose levels. The probit and log probit models are common tools employed to fit such sigmoid shape dose-response curves and for calculating confidence intervals for dose-response quantiles, such as ED_{01}. The probit model takes the form

$$\Pr(r = 1 \mid d) = \Phi(d'\beta),$$

where tumour response r takes on a binary form of 0 (absence) and 1 (presence), D is the dose `regressor' that influences outcome r, and Φ is the cumulative distribution function of the standard normal distribution. The β parameter is estimated, and thereby the curve is fitted, using maximum likelihood estimation.

An equivalent yet simpler form to interpret the probit model is:

$$Y' = \Phi^{-1}(p),$$

where Y' is the probit transformed value, p is the proportion of tumours (i.e. number of tumours observed/total number of animals observed) at a given dose-level or time (depending on the relationship being considered). The link function, $F(p) = \Phi^{-1}(p)$, takes the form of the cumulative normal distribution and is known as the probit link, short for probability unit. The inverse $F(p)$ is the $100 \times p\%$ quantile of the standard normal distribution.

Mantel & Bryan have noted that doses that produce carcinogenic responses appear to be lognormally distributed, i.e. plots of tumour incidence as probits versus log dose were approximately linear, and that the slopes were steep and generally greater than one [14, 15]. The log-probit model uses a log link function. Suppose the tumour occurs with probability p then the odds of it happening are $p/(1-p)$. The odds ratio is non-negative, taking logs, the logit link (or log odds) is $\log(p/(1-p))$. The associated density function is similar to the standard normal distribution, with thinner tails. It was proposed [17] that an upper bound on the dose associated with some *a priori* acceptable level of cancer risk could be estimated by extrapolating from an upper confidence limit on the estimate of risk at the lowest experimental dose, and modified by [18] to consider a minimum excess risk of 1%.

One of the criticisms highlighted for the Mantel-Bryan model is that although intended to provide conservative risk estimates, extrapolation in the low dose region often produces risk estimates that are higher than those estimated using other dose response techniques [19]. Hogan [19] also states that there is no mechanistic model for carcinogenesis that is reasonably approximated by a log-probit distribution model.

iii. The Hit Models

The third class of dose response models, referred to as the "hit" models, are so-called because they are intended to take into consideration the hypothesised mechanistic aspects of tumour formation. The simplest of the models within this category is the one-hit model, based on the biological principle of the single mutation-single cell hypothesis for carcinogenesis [16, 19]. This model takes the form

$$P(d) = 1 - \exp\{-\lambda d\},$$

Where λ is the unknown model parameter to be estimated and λ*d* is the expected number of hits at dose level *d*. Use of this model is reported to produce results similar to those obtained using the linear model. However, one of advantages over the linear approach is the ability to use all of the experimental data to estimate the model parameters [19].

The dose-response model, perhaps most commonly used for quantal data (data indicating only number of animals with cancer), is the multi-stage model. The model accounts for the bio-mechanistic process by assuming that a cell goes through a number of distinct stages (*k*) before becoming malignant. The underlying basis is that tumour incidence will increase as a function of age. Such models have been adapted to, for example, include the effect of exposure to a carcinogen by assuming that the transition rate at which a cell goes through (at each stage) is linearly related to the dose rate, that is dose *d* is a dose rate of a continuously administered carcinogen. The (linear) multi-stage model takes the form

$$P(d) = 1 - \exp\{-(\lambda_0 + \lambda_1 d + \lambda_2 d^2 + \dots \lambda_k d^k),$$

where λ_i (the linear terms) are unknown non-negative parameters to be estimated (using maximum likelihood methods), *d* is the average lifetime daily dose of the chemical (often in mg/kg/day), and *P(d)* is the lifetime probability of cancer from dose *d*. In addition, an upper confidence limit on the dose response curve is also calculated, reflecting the uncertainty of extrapolating the curve to low doses [20].

For risk assessment, the interest lies in the extra lifetime risk of cancer resulting from exposure to the carcinogen at dose d, and hence the total risk can be expressed as

$$P*(d) = P(0) + [1 - P(0)P(d)],$$

where the risk (or excess risk) due to exposure is

$$P(d) = \frac{P*(d) - P(0)}{1 - P(0)},$$

and can be interpreted as the probability of occurrence of a tumour at dose *d*, given that no tumour would have occurred in the absence of a dose [20].

Multi-stage or hit models in their various forms have perhaps been the most widely employed, nevertheless, notwithstanding criticisms. While behaving like a simple linear model, it assumes that background tumour incidence and that due to a carcinogenic chemical are additive, which may be an invalid assumption. Another form of the multi-stage model, taking the form of a Gamma distribution, treats the tumour rate as independent of background which is difficult to verify empirically and can have tremendous impact on magnitude of risk estimates [19] .

II. Evidence concerning the empirical form of dose-response relationships at low dose exposures.

The margin of exposure has no units and hence, unlike the VSD, superficially it does not imply any specific level of risk. However, in concluding that MOEs of >10,000 are of low concern [21, 22], it is at least implicit that the risks associated with such exposures are tolerable. There have been suggestions that a value of at least 10,000 is necessary to take account of a number of uncertainties and variability between and within species. However, in general the sources of such uncertainty and variability have not been well defined. It has been argued by some that MOEs greater than 10,000 are of low concern because the risk associated with such exposures, assuming linear extrapolation, would be at a level where there was a low probability of harm. When the MOE is used to rank chemicals for prioritization, for example for remediation, there will be some implicit assumption about the dose-response relationship underlying different MOEs. The importance of the difference between MOEs will be dictated by the steepness of the respective dose response curves. Interpretation of the level of concern associated with the MOE would be considerably improved if the nature of the dose-response curve for genotoxic carcinogens could be substantiated by sound scientific considerations.

As discussed earlier, the mathematical models that are used to estimate the risks from low dose exposures differ substantially. The different models used come with their own limitations and the risk estimations differ immensely. The aim of the present study was to carry out an exhaustive review of evidence that relates to the different mathematical models that have been used to describe the empirical form of the dose response curve for genotoxic carcinogens. This was done to determine whether any tendencies exist for such chemicals to follow a particular form (e.g. linear, multistage etc) and whether these tendencies apply to genotoxic carcinogens in general or vary according to the different classes of genotoxic carcinogens.

2.1 Collation of relevant evidence

A search strategy was developed to identify published studies that have sought to investigate the empirical form of the dose response curve at low levels of exposure of experimental animals to genotoxic chemicals. The search was carried out using the search engine PubMed followed by citation-searching within the references of relevant publications to identify any other potentially relevant publications. Experts were also consulted for relevant papers. The reader is referred to Appendix II for a flow diagram describing the search strategy implemented and the numbers of papers shortlisted as potentially relevant/ relevant as part of the search process.

The papers were initially selected on the basis of key words within the titles of the published papers. The key words included: genotoxic carcinogens, low dose exposure, dose response, linear, nonlinear, etc. The relevance of the selected studies to the review topic was then examined further by either reading the abstract or where necessary the entire paper.

Although the initial search string compiled identified relevant papers, other relevant papers were also identified on an *ad hoc* basis. The latter highlighted the restrictive nature of the initial search strings that were used in PubMed and prompted the need to refine the search strategy (outlined below).

Table 1: Initial search results

Search 1:

Search number	Search criteria	Hits	Potentially relevant	UNIQUE (i.e. different hits from search no. 1
1	"threshold" AND "genotoxic*"	263	35	[reference set (35)
2	"threshold" AND "genotoxicity"	139	14	3
3	"low dose" AND "threshold" AND "genotoxic*"	28	19	2
4	"threshold" AND "genotoxic" AND "carcinogen"	31		0
5	"low dose" AND "dose response" AND "genotoxic" AND "carcinogen"	16	4	3
6	"low dose" AND "dose response" AND "genotoxic*"	108	29	13
7	("dose response" AND ("genotoxic" OR "genotoxicity" OR "genotoxins") AND ("Carcinogen*")	262	9	4

Search 2:

- Key words were compiled from a number of identified key papers in order to design a more comprehensive search strategy.

((genotoxic* OR carcinogen* OR cancer OR mutagen* OR tumor OR preneoplastic OR "DNA adducts") AND (Threshold OR "linear no-threshold" OR extrapolation OR "dose response" OR "dose response assessment" OR linear) AND ("linear extrapolation" OR "low dose" OR slope OR (modeling OR modelling) OR "risk assessment")) =3337

Limits in PubMed: studies in animals and publications in English only

Hits: 3337

Potentially relevant: **46** (this is inclusive of papers identified by citation searching)

Search 3: Revised search string

Although this search identified a significant number of papers it was noted that some key papers (identified from search 1) were still not being picked up by the search. The search string was revised to:

((genotoxic* OR carcinogen* OR cancer OR mutagen* OR tumor OR preneoplastic OR "DNA adducts") AND (Threshold* OR "linear no-threshold" OR extrapolation OR "dose response" OR "dose response assessment" OR linear) AND (nonlinear OR "non-linear" OR "linear extrapolation" OR "low dose" OR slope OR (modeling OR modelling) OR "risk assessment"))

Limits in PubMed: studies in animals and publications in English only

Hits: 3717

Potentially relevant (different from search 2): **6**

Expert solicitation

One of the main problems with identifying relevant papers within this subject area is the broad range and inconsistent use of key words that are currently used to tag relevant papers in PubMed. Although the refined search was successful in identifying more relevant papers, it was concluded that it may not be feasible to identify all of the relevant papers using such a systematic approach. Help was therefore solicited from known experts within the field of low dose cancer risk assessment (identified from the participants of the elicitation workshop carried as part of objective 03/01 of this project).

In addition to the solicitation of relevant publications from experts within the field, the criteria for the selection of relevant papers were also refined to the following:

- Studies should provide experimental evidence for the nature of the dose-response at low doses.
- Studies should present new data, and not be a repeat of published information.
- Preference should be given to studies that have investigated tumourigenic endpoints. Studies should be on endpoints relevant to genotoxic carcinogenicity.

Table 2: Summary of identified papers

Searches	Shortlist of potentially relevant papers identified from searches	Potentially relevant papers
Search 1	60	11
Search 2	3337	46 (also identified all papers from search 1).
Search 3	3717	6 (unique from those identified by search 2)
Expert solicitation	51	7 (unique from searches 1 -3)

A thorough assessment of all "potentially relevant" papers from the searches and expert solicitation identified **17** relevant papers (summarised in table 3). Ten different genotoxic chemicals, which fell under the categories of group 1 and group 2 carcinogens as classified by IARC[1], were investigated in the studies summarised. The study investigators employed a range of animal models that included rats, mice and trout. The only *in vitro* studies encountered employed human and Chinese hamster lymphoblastoid cell lines.

International Agency for Research on Cancer (IARC) – carcinogen classification scheme:
[1] **Group 1**: Carcinogenic to humans.
Group 2A: Probably carcinogenic to humans.
Group 2B: Possibly carcinogenic to humans.
Group 3: Unclassifiable as to carcinogenicity in humans.
Group 4: Probably not carcinogenic to humans.

Table 3: Summary of relevant studies (17)

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
National Centre for Toxicological Research (NCTR) [1]	2-acetylaminofluorene (2-AAF) administered in the diet. ED01 study *In vivo*	Diet	24,192 female BALB/c, StCrlf, C3Hf/Nctr mice	0, 30, 35, 45, 60, 75, 100 and 150 ppm	Liver and bladder cancer	Probit log-dose model/ one hit linear	Significant incidences of liver and bladder tumours reported. Incidences of bladder cancer reported to decline sharply as the dosage of 2-AAF was reduced. Liver cancer reported to show a nearly linear response over the experimental dose range studied.
Gaylor DW [re-evaluation] [23]	2-acetylaminofluorene (2-AAF) administered in the diet. ED01 study *In vivo* study	Diet	20880 female BALB/c, StCrlf, C3Hf/Nctr mice *In vivo*	0, 30, 35, 45, 60, 75, 100 and 150 ppm	Bladder cancer	Probit log-dose model/ one hit linear	Re analysis of the ED01 data available for bladder tumour by the original study investigators to clarify the shape of the d-r curve at the low dose range for the observed incidences of urinary bladder neoplasms. Focus on definition and thorough classification of tumours. Observed tumours examined by a single pathologist. Original d-r curve for bladder tumours reported as flat response at low dose followed by sharp rise at higher doses which gave the impression of threshold. Only Grade A & 0 tumours were re-examined by one pathologist. Some difficulty encountered in distinguishing grade 0 bladder carcinoma, so they were excluded from reanalysis. The examined grade A-D tumours did not show a low dose trend from 0-60 ppm of 2-AAF. Authors concluded that the nature of d-r below 60 ppm remains uncertain.

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
Society of Toxicology [24] [re evaluation]	2-acetylaminofluorene (2-AAF) administered in the diet. ED01 study *In vivo* study	Diet	20880 female BALB/c, StCrlf, C3Hf/Nctr mice *In vivo*	0, 30, 35, 45, 60, 75, 100 and 150 ppm	Bladder cancer	Hattley-Sielken time to tumour model	A task force set up by the SOT reanalysed the ED01 data after concluding that the initial NCTR analysis did not account for time to tumour analysis in the modelling of the data. The SOT applied the Hartley-Sielken model, and observed that the response observed was non linear. They concluded that a thresholded effect was clearly shown for bladder tumours below 60 ppm).
Zapponi GA [25]	Vinyl Chloride In vivo	Inhalation	Rat (S.D, Wistar). Mouse (CD1, swiss)	50-2500 ppm (taken from 7 different studies	Liver angiosarcomas	Weibull, multistage and linearised multistage models	The authors reviewed results produced from 7 different experiments that employed different doses and 2 different species and 2 different strains in their different studies (to evaluate consistency among results). The authors reported a supralinear trend for VC in range of experimental doses. Downward trend observed was interpreted as being inconsistent with the hypothesis of a threshold. Different species and strains showed highly consistent results.
Peto R [26]	N-Nitrosodiethylamine (NDEA) & N-Nitrosodimethylamine (NDMA). In vivo	Diet (drinking water)	4080 (2040 males; 2040 females) Inbred Colworth rats	16 doses (0 to 16.896 ppm -lowest dose tested was 0.033 ppm)	Liver and oesophagus tumours	Weibull distribution	Linear relationship at low dose rates (below 1 ppm) reported. Authors suggested that a dose of 1 ppm in drinking water, of NDEA or NDMA would result in a 25% excess of tumours.

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
Poirier MC [27]	2-AAF	Diet	BALB/c mice (4-5 animals per group)	0, 5, 10, 15, 30, 45, 60, 75, or 150 mg/AAF kg diet	DNA adducts	Weibull analysis	Experiment carried out to correlate the formation of DNA adducts at low dose exposure of mice to results of tumour formation obtained from ED01 study. Linear correlation observed between administered doses of 2-AAF and concentration of DNA adducts found in both bladder and liver. However D-R relationship between tumour induction in liver and bladder differed between the tissues. Conclusion: DNA adducts are necessary but not sufficient for tumourigenesis.
Purchse & Auton [2] [re-evaluation]	2-AAF	Diet	24,192 female BALB/c, StCrlf, C3Hf/Nctr mice	0, 30, 35, 45, 60, 75, 100 and 150 ppm	Liver and bladder cancer	Log/linear and log/log plot	A review of the ED01 data and general discussion of thresholds in carcinogenesis. The ED01 data for both bladder and liver cancer was re-plotted on a log/linear and log/log plot.

The authors reported that the log/linear plot was shown to give an even stronger appearance of a threshold for bladder cancer.

Whereas the log/log plot of the data showed no threshold for bladder cancer. |

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
Williams GM [28] (part of a series of paper published by the same author for the same compound –1993, 1999, 2000, 2004)	Diethylnitrosamine (DEN) In vivo	Diet	108 (male) F344 rats	1, 2 or 4 mmol/kg body weight	AHF & GS+	No formal modelling or statistical assessment was carried out on the dose response data.	Quantification of hepatocarcinogenesis through the identification of different preneoplastic lesions (altered foci) and histochemical markers in the liver: -level of glutamine synthetase (GS+) in hepatocytes. Nonlinear response reported on the basis of the magnitude of multiplicity of foci at the different dose levels.
Choy WN [29]	Aflatoxin B1 (AF) In vivo	IP & diet (various studies compiled)	Rat, mice & monkey	IP study - 10, 25, 65, 160, 390, 1000 ng/kg; Diet study: 1, 10, 102, 103, 104, 105 & 2x105 ng/kg Liver tumour study: 5 doses, lowest dose- 54 ng/kg/d	DNA adducts (liver)	Linear function fitted to the log-transformed data. Linearized multistage model used for low D-R analysis of liver tumour induction.	IP study (DNA adduct formation): D-R based on only one of the 13 reviewed studies. Dose response analysis using linear model fitted data (R-Square = 0.9685/ adjusted = 0.9607) slope = 1.07; SE 0.096. Ingestion studies: composite D-R curve based on 3 different rat studies. Results showed linear model fitted perfectly (R-Square =0.9956/ adjusted = 0.995; slope 1.01 SE 0.024; hypothesis accepted. No threshold observed from either ingestion or IP studies at the doses tested. Liver tumour induction shown using data from 1 out of the 7 studies reviewed. D-R analysis carried out using linearized multistage model - plot showed response was linear without a threshold. The authors opted not to comment on the validity of the assumptions inherent in the model used.

15

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
Shirai T [30]	2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) In vivo	Diet	Sprague-Dawley (SD) & F344 rats	0, 25, 100 or 200ppm	Tumours of colon & mammary	Weibul and one hit model	Medium and long term tests carried out in rats. D-R study: authors described a clear d-r relationships described for both latent tumour appearance and final tumour incidence. Although indicated that the Weibull and one hit (poor fit) models were used, a dose response graph was not shown. Instead a plot of % of tumour against time was shown. VSD at risk level of 10^{-6} calculated using Weibull model (0.023-0.52ppm in diet) which was higher than the daily intake in humans.
Williams GM [31] (part of a series of paper published by the same author for the same compound –[28, 31-33]	DEN In vivo	IG	390 F344 rats	25, 50, 100 and 200 µmol/kg/bodyweight	DNA adduct formation, hepatocellular alered foci & liver cell proliferation	Least squares regression Significance at $p < 0.05$.	Observations for effects at the low exposures were taken at 5 weeks. Non linear relationships observed for early DEN effects (DNA adduct formation, hepatocellular altered foci & liver cell proliferation).

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
Williams GM [33] (part of a series of paper published by the same author for the same compound –[28, 31-33]	2-AAF & DEN	IG	60 F344 rats	**DEN**: LE = 0.5 mol/kg/bw; MLE=1 mmol/kg/bw ; LE = 2 mmol/kg/bw ; MHE= 3 mmol/kg/bw ; HE = 4 mmol/kg/bw ; **AAF**: LE = 0.5mmol/kg body weight	DNA adducts and tumours	No statistical modelling was described by the authors. The authors however described the observed endpoints in terms of low exposure (LE) and high exposure (HE).	Tumours were only observed in the highest cumulative dose tested and not in the low exposures. The authors reported that the formation of DNA adducts can be non linear, with a plateau at higher doses; no effect levels were observed for hepatocyte proliferation. The authors reported that the findings were indicative of thresholds for the hepatocellular initiating of DEN and AAF. Low exposure (LE), Mid-low exposure (ME), Medium exposure (ME), Mid-high exposure (MHE), High exposure (HE).
Waddell WJ [34] [re-evaluation]	2-acetylaminofluorene (2-AAF) In vivo	Diet	21,192 female BALB/c, StCrlf, C3Hf/Nctr mice	$0, 10^{19.10}, 10^{19.17}, 10^{19.28}, 10^{19.40}, 10^{19.50}, 10^{19.62}, 10^{19.80}$ molecules/kg/day	Bladder and liver tumours	Re analysis of Ed01 data. The data was plotted on a logarithmic scale using a dose scale proposed by Rozman et al (1996) (referred to as	Tumour response was plotted linearly against logarithmic dose (0-100% tumours) to facilitate the calculation of threshold dose at zero tumours. The authors converted the doses administered into molecules/kg/day. A clear and consistent threshold was reported for both bladder and liver neoplasms at about 10(e19.5) and 10(e19.1) molecules/kg/day.

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
						the Rozman scale).	
Waddell WJ [35] [re-evaluation]	N-Nitrosodiethylamine (NDEA) & N-Nitrosodimethylamine (NDMA). In vivo	Diet (drinking water)	4080 (2040 males; 2040 females) inbred Colworth rats	16 doses (0 to 16.896 ppm -lowest dose tested was 0.033 ppm) - doses were converted to molecules/kg/day	Liver and oesophagus tumours	Re-analysis of study carried out by Peto et al 1991. The data was plotted on a logarithmic scale using the rozman scale.	The author considered that the inability of Peto et al [26] to reliably predict the shape of the dose response data in their initial study could be resolved by the reanalysis of the data using a logarithmic scale for dose. The author reported that there was a convincing sharp threshold at $10^{17.1}$ molecules/kg/day for the effect of NDEA on oesophageal cancer.
Williams GM [32] (part of a series of paper published by the same author for the same compound –[28, 31-33]	2-AAF In vivo	IG	60? (not clear although control group was 60 F344 rats	112 – 448 mg/kg per body weight	Preneoplastic effects	No statistical or modelling methodology was described by the authors. Tumour incidences obtained for cumulative high, medium and low doses (CHD, CMD and CLD). The effects were	The authors reported practical thresholds for all 4 of the effects examined (formation of DNA adducts, arylsulfotransferase & glutamine-synthetase (GS) zone, replicating fraction (RF) and hepatocellular altered foci (HAF) to reflect the hepatocellular initiating effects of AAF. The formation of DNA adducts was described as nonlinear and the effects on HAF were described as being supralinear at CHD.

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
						described as either linear, nonlinear or supralinear.	
Doak SH [36] Gocke E [37]	EMS and ENU	Gavage	Out-bred Crl: CD-1 (ICR) Mice	EMS: 1.25 to 260 mg/kg/day (low doses 0, 1.25, 2.5 5 mg/kg/day) ENU-1.1 to 22 mg/kg/day (low doses (0, 1.11, 4.45mg/kg/d ay)	Micronucleus induction	Four step statistical assessment: -Comparison of control groups -Rejection of linear dose-response relationship (entire dose range). -acceptance of linear dose-response relationship below the NOEL.	In vivo MNT: the authors reported a clear threshold in the dose-response for EMS. Doses of up to 80 mg/kg/d did not result in the induction of MN. Cytotoxic effects were observed only at the highest dose of 260 mg/kg/d. [Study included because it investigated tumourigenic effects as part of the study – even though effects were only observed at the highest doses tested].

First author Year	Carcinogen(s) tesed/ type of study	Route of admin	Species & no. of samples	Doses investigated	Parameter/ Endpoint investigated	Statistical/ modelling method	Observations/ comments
Bailey GS [38]	Dibenzo[a,l]pyrene (DBP) In vivo	Diet	40,800 Rainbow trout (Oncorhyncus mykiss	0 – 225 ppm (low doses tested: 0, 0.45, 1.27, 3.57,10.1 ppm)	Stomach and liver cancer. DNA adducts.	Nine different models explored. Only 3 fit the data: linear probit, quadratic logit and Ryzin-Rai.	The best fit models were compared against the default linear extrapolation model (LED10) to calculate the level of conservativeness. A non linear (sub linear) response for both liver and stomach cancer was reported. Estimation of the virtually safe dose (VSD) using the LED10 model showed that the risk estimation was 500-1500 times more conservative than using other models.

III. Overview of studies designed to examine the dose response curve for genotoxic chemicals at low dose exposures.

As discussed earlier, one of the main problems with carrying out studies in animal models is the extrapolation that is required from animals to humans and also the extrapolation of information from the dose response curve at the high doses normally and necessarily used in the animal studies to the low doses to which humans are generally exposed in order to carry out human risk estimations.

Various methods and mathematical models were applied by the different authors to investigate the shape of the dose response curve at very low levels of exposure. The authors based the dose response analysis on either tumour or preneoplastic endpoints. In the case of some chemicals (e.g. 2-AAF and DBP), the dose response relationship was investigated for both endpoints. The various studies are discussed in more details in the following sections.

3.1 The Mega Studies

In an effort to address the issue of carcinogenicity testing at dose levels similar to human exposure levels, the so called mega studies were conducted. The mega studies required the use of extremely large numbers of animals in an effort to reduce the sampling error to a low enough level to enable very small effects to be detected [12, 39].

The search strategy employed here identified three such studies, carried out in the mouse, rat and trout; these are discussed in more detail below.

3.1.1 The Mega-Mouse/ED01study

An experiment conducted by the National Centre for Toxicological Research (NCTR) saw 24,192 female BALB/c StCrlf C3Hf/Nctr mice administered the carcinogen 2-acetylaminofluorene (2-AAF) allocated to 81 different treatment groups [39, 40]. The mega-mouse study was also termed the ED01 study, so-called because it was designed to estimate with precision the effective dose that produced a one percent increase in tumour incidence [40]. The study investigators chose the one percent tumour incidence as a response as this provided the opportunity to study the shape of the dose response curve with one order of magnitude greater precision than the level normally obtained from general carcinogenesis screening studies carried out at the time the study was conducted.

The ED01 study was not set up with the intention of finding a threshold for a genotoxic carcinogen, instead, the study was designed to estimate the dosage that produced a 1%

tumour incidence detectable above the spontaneous background rate [39, 41]. This was in contrast to normal studies that produce tumour incidence rates at 4-6% detectable above the spontaneous background rate [5]. The mice were administered doses of 0, 30, 35, 45, 60, 75, 100 and 150 ppm 2-AAF in the diet. The study investigators reported that significant levels of tumours were observed for bladder and liver cancer among the mice investigated. A summary produced by the NCTR [39] reported a nearly linear response over the experimental dose range leading the study investigators to dismiss the notion of a threshold dose. The incidence of bladder tumours was observed to decline in response to the reduction in dose of 2-AAF[39].

The study investigators applied the probit method described by Finney [42] to analyse the data for both dose response and time response. This model was considered to adequately describe the relationship of cancer endpoints for both dose and time. In addition the model was used to aid the biological interpretation of the results [43].

A total of 20346 and 20428 mice for bladder and liver neoplasms respectively were used to model the experimental data [43]. Owing to the low incidence of liver and bladder neoplasms observed at all but the high doses, the study investigators made no attempt to fit a dose response prior to 18 months [43].

The study investigators modelled the data for each time interval as well as for each dose level. An attempt was also made to model the incidence data and relate it to both time and dose. The study investigators reported that although attempts were made to fit the probit log time model to all available data, they did not observe sufficient incidence of bladder tumours below 60 ppm so they were unable to fit the model to these data. It was noted that the probit log-time regression line for the treated animals was not different from that observed for the controls. Figure 1 below shows the dose response relationships obtained from the NCTR analysis. Both dose response curves for liver and bladder tumours for data between 18 and 33 months were shown to have very steep slopes.

Figure 1: The prevalence of bladder neoplasms in sacrificed mice in respect to time on study (obtained from Littlefield et al [1])
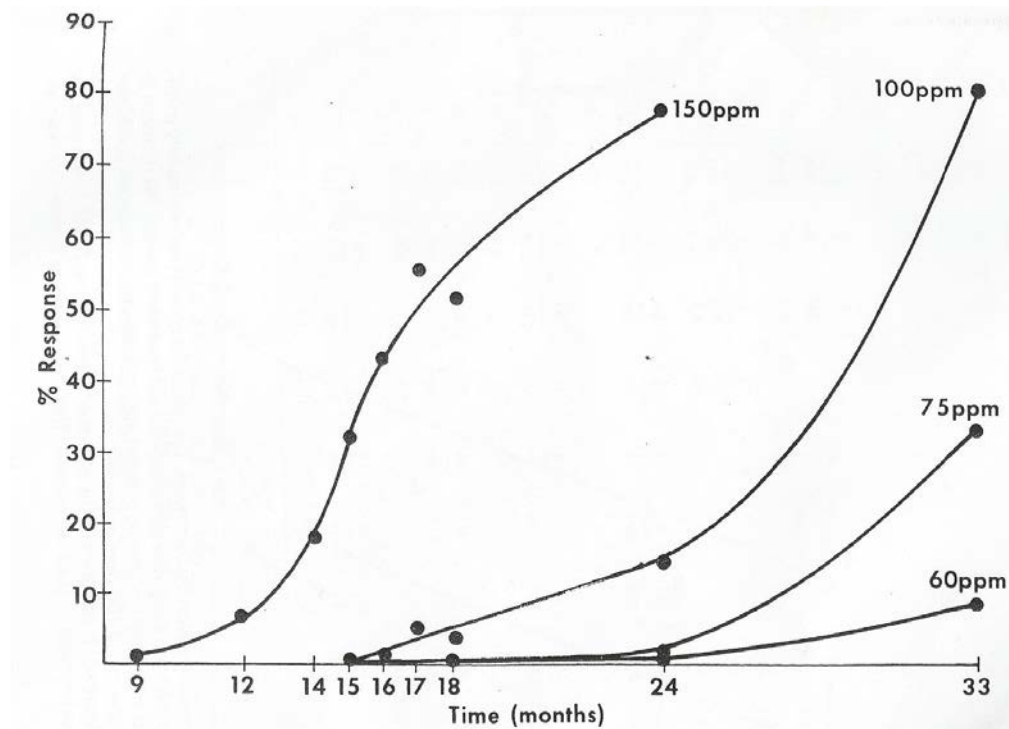

Figure 2: The prevalence of liver neoplasms in sacrificed mice in respect to time on study (obtained from Littlefield et al., [1])
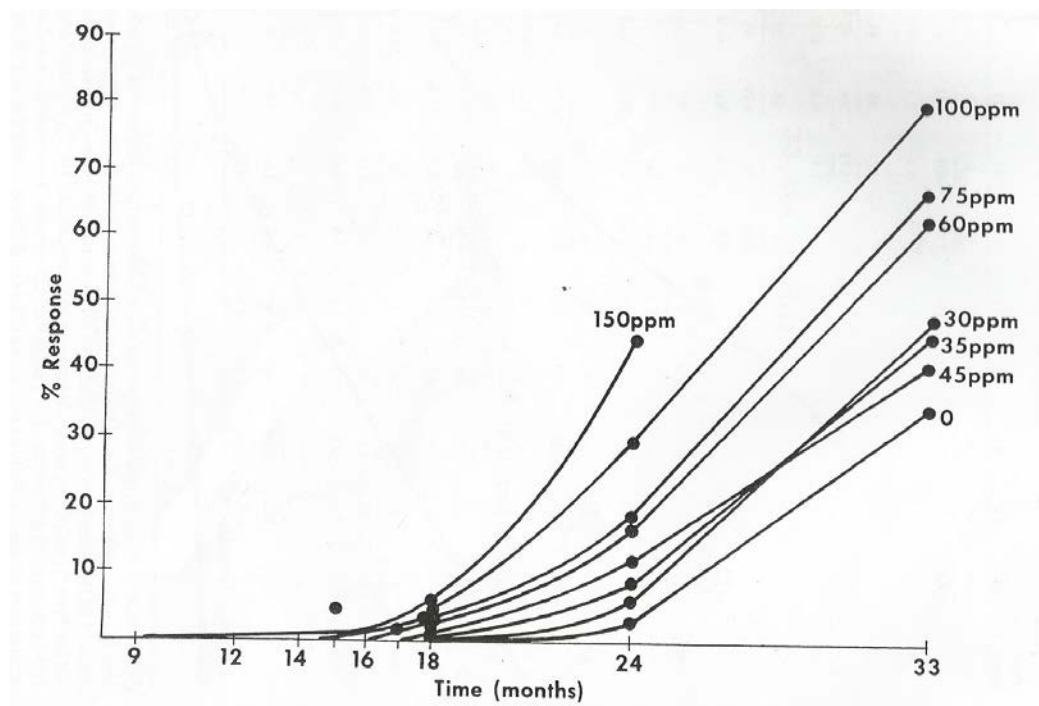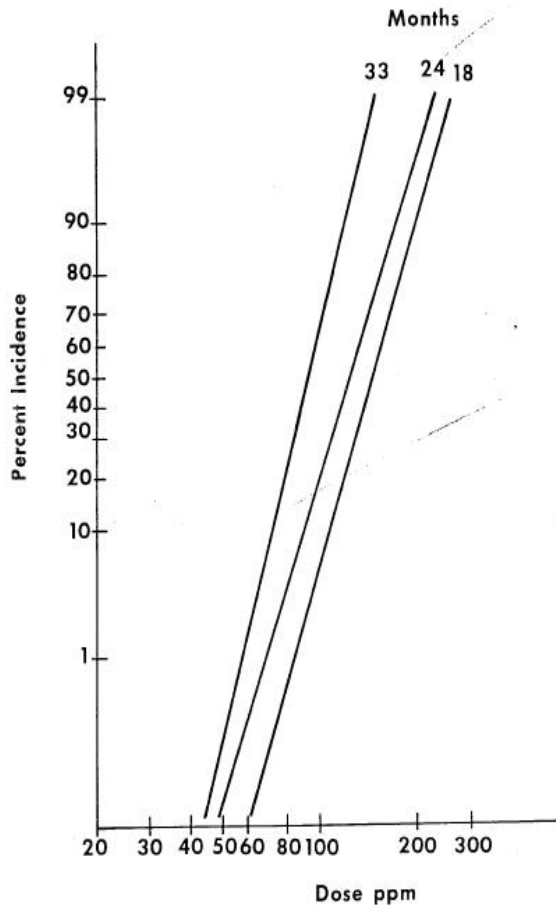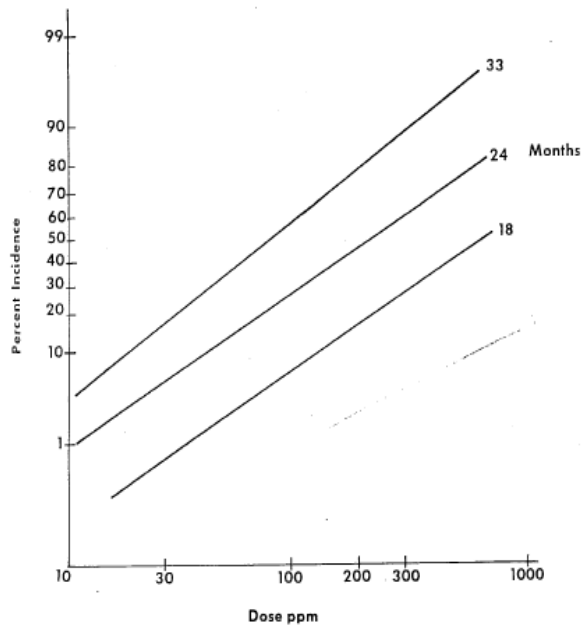
Figure 3: Probit log dose models for bladder and liver neoplasms

(obtained from Farmer et al [44])

Probit log dose model for bladder neoplasms

Probit log dose model for liver neoplasms

3.1.2   Re-examination of ED01 study results

Society of Toxicology

The results produced by the NCTR were widely anticipated by the scientific community because they were from such a large scale study that examined the effects of a well known genotoxic carcinogen at low doses of exposure. Upon publication of the results from the ED01 study, disappointment was registered by the Society of Toxicology (SOT) at some of the conclusions reached by the NCTR in their published summaries and conclusions as it was considered that they were more by "politico-regulatory" in nature than scientific [24].

The SOT council set up a committee that included toxicologists, pathologists and statisticians to re-examine the ED01 data. One of the main concerns highlighted by the SOT task force was the application of the statistical model used to analyse the ED01 data. It was the opinion of the task force that traditional time-to-tumour models such as the one-hit model, two-parameter Weibull model and two-parameter extreme value model did not fit the data generated from the ED01 study and were too simplistic and inflexible for such a dataset. It was considered that the inappropriate use of the one-hit model employed in the ED01 analysis was best highlighted by the lack of fit of the model for the bladder tumour data.

The SOT re-examination of the data focused on the re-analysis of the functional relationship of the dose and time data to the carcinogenic response. The task force concluded that the initial data analysis by NCTR did not account adequately for time-to-tumour and their estimation of a virtually safe dose on the basis of only quantal response data was considered to be insufficient because such risk assessment requires the additional consideration of the time-to-tumour data [24, 45-47].

The task force suggested that the Hartley-Sielken model, based on the multistage theory of carcinogenesis, was a better fit for the ED01 data. This model was considered to fit both the liver and bladder tumour data well and was reported to model both time-to-tumour and dose factors and was more appropriate for a dataset of the ED01 magnitude. They concluded that when the bladder tumour data were analysed by time-to-tumour techniques the shape of the dose response curve was clearly not linear as initially described by the NCTR. More importantly the SOT task force went even further to state that below the treatment level of 60 ppm AAF in the diet, it was clear that AAF had no effect on the induction of bladder tumours. Re-modelling of the bladder tumour data showed that indeed the shape was non-linear and there appeared to be a threshold at this dose level. It was noted that when the time-to-tumour factor was not incorporated in the analysis, the shape of the curve appeared to be linear [24].

Purchase & Auton [2]

A review carried out by Purchase and Auton to discuss the issues surrounding thresholds in carcinogenesis involved a re-analysis of the ED01 data. In acknowledgement of the controversy surrounding the issue of a threshold for the ED01 data, particularly for bladder tumours, the authors re-plotted the datasets available. The same data plotted on a log/log and log/linear scale produced very different results. The log/linear plot produced an even stronger appearance of a threshold while with the log/log plot a threshold was less evident.

In addition the authors calculated the smallest group size necessary to detect a statistically significant change in incidence of the same size as that observed, assuming the control rate is known exactly. The number of animals required for this for bladder cancer was 27,000 in the lowest dose group (30ppm). This value was in contrast to the actual numbers used in the ED01 study where the study investigators employed replicates of between 1 and 2000 (experimental room was designated a replicate). Clearly it would not have been feasible to have individual group sizes of up to 27,000.

Figure 4: Incidence of liver and bladder tumours after 48 months in the
Ed01 study (Littlefield et al. [1]) – log/ linear scale
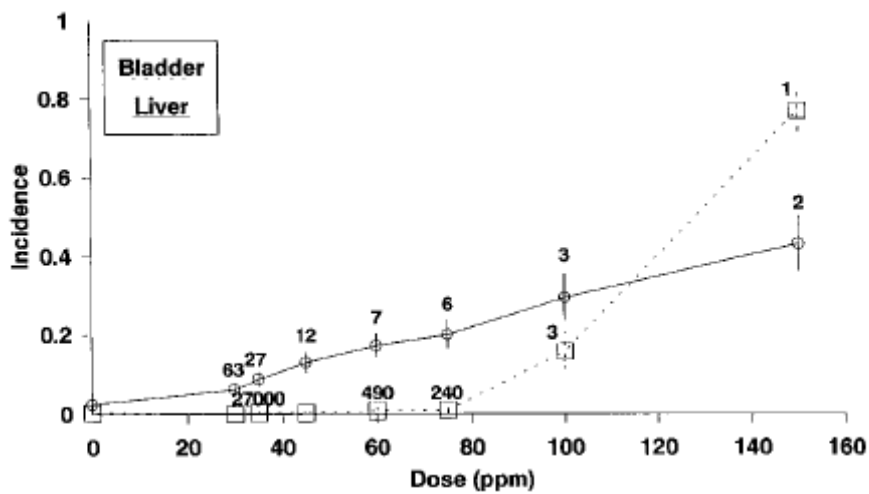
Source: Purchase & Auton [2]

Figure 5: Incidence of liver and bladder tumours after 48 months in the ED01 study (Littlefield et al., [1]) – log/logscale.
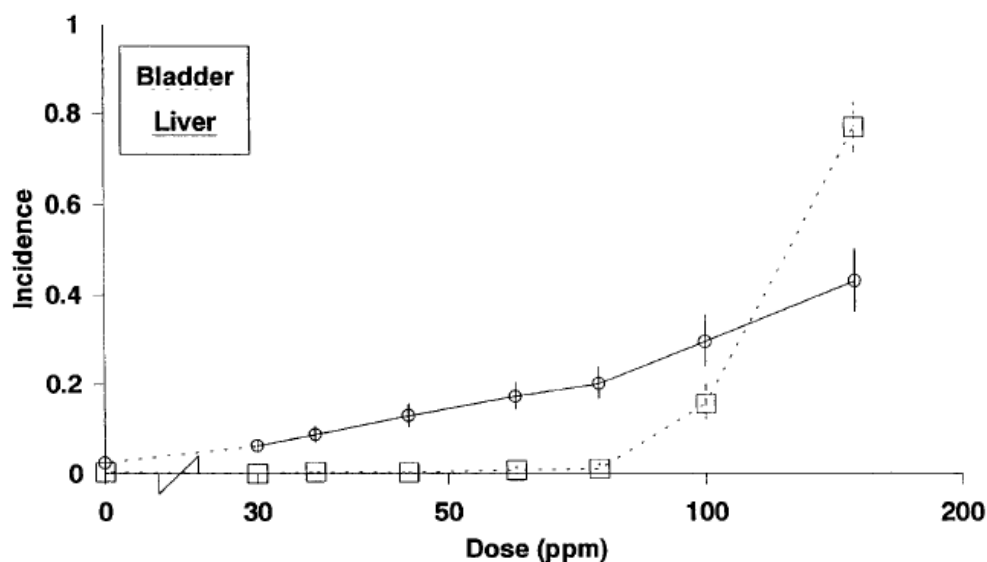
Source: Purchase & Auton [2]

Waddell [34]

The ED01 data were also re-analysed by Waddell [34] as he also criticised the conclusions and modelling techniques applied by the NCTR. It was the opinion of Waddell that representation of the ED01 data using a procedure proposed in his earlier publication [48] using a dose scale proposed by Rozman et al [49] (referred to as the Rozman scale) was a much more appropriate way to represent the observed results. This was for the following reasons:

- Plotting of the dose on a logarithmic scale is considered to best mimic biological responses, thought to be related to the logarithm of the dose.

- The Rozman scale uses number of molecules instead of weight as a measure of the dose of the chemical. This was thought to be advantageous particularly for the comparison of compounds[2] with varying weights.

---

[2] It should be noted however that this was not advantageous in the case of the ED01 data as only one compound was investigated.

- The scale enables the consideration of a very wide range of doses, right down to one molecule (10^0). This was considered to be advantageous particularly when considering the notion that a single molecule might be capable of causing cancer.

It was the opinion of Waddell that the historically acknowledged uncertainties associated with dose response assessments, particularly over whether a threshold can be observed for animals exposed to carcinogens, was nothing more than an error in the plotting of the dose response [34].

Upon re-evaluation of the ED01 data Waddell concluded that thresholds were observed for both liver and bladder cancer. Waddell [34] reported a clear and consistent threshold for both bladder and liver neoplasms at $10^{19.5}$ and $10^{19.1}$ molecules/kg/day respectively (see figures 6 to 7). Waddell [34] observed very different slope shapes for the two endpoints. Firstly the slope of the dose-response curve for bladder neoplasms was observed to be very steep from 17 through to 33 months whereas the slope for liver neoplasms was observed to increase from a shallow slope at 18 months to a steep slope at 33 months. The author concluded that this difference in slope shape between the two endpoints was most likely to have been the result of the different mechanisms of carcinogenesis within the two organs.

Figure 6: Figure obtained from Waddell [34] shows the percentage of all mice (dead, moribund, and terminated) at 24 months, with bladder neoplasms plotted against molecules of 2-AAF/kg/day. Waddell [34] obtained the data from Table 1 of Farmer et al [43].
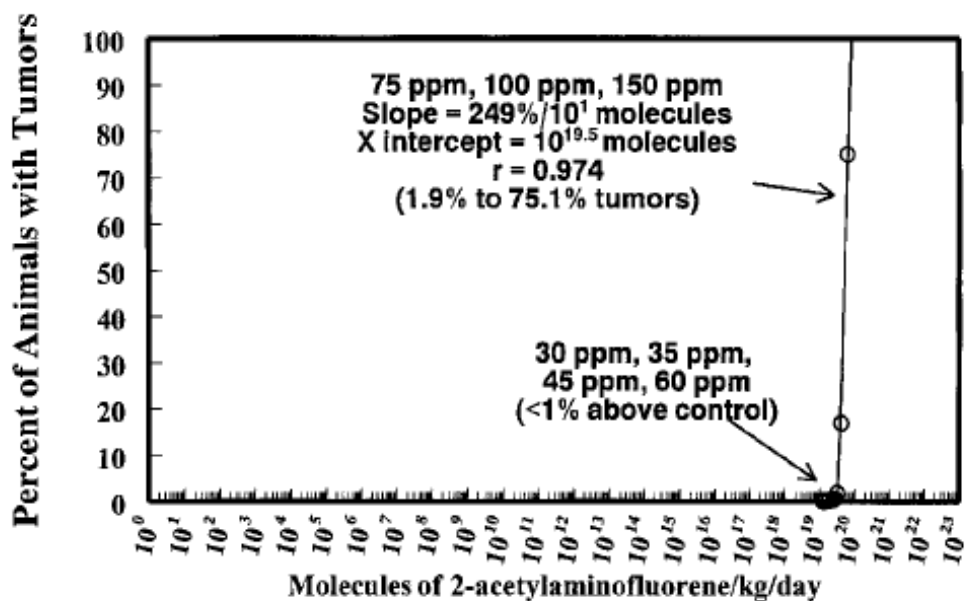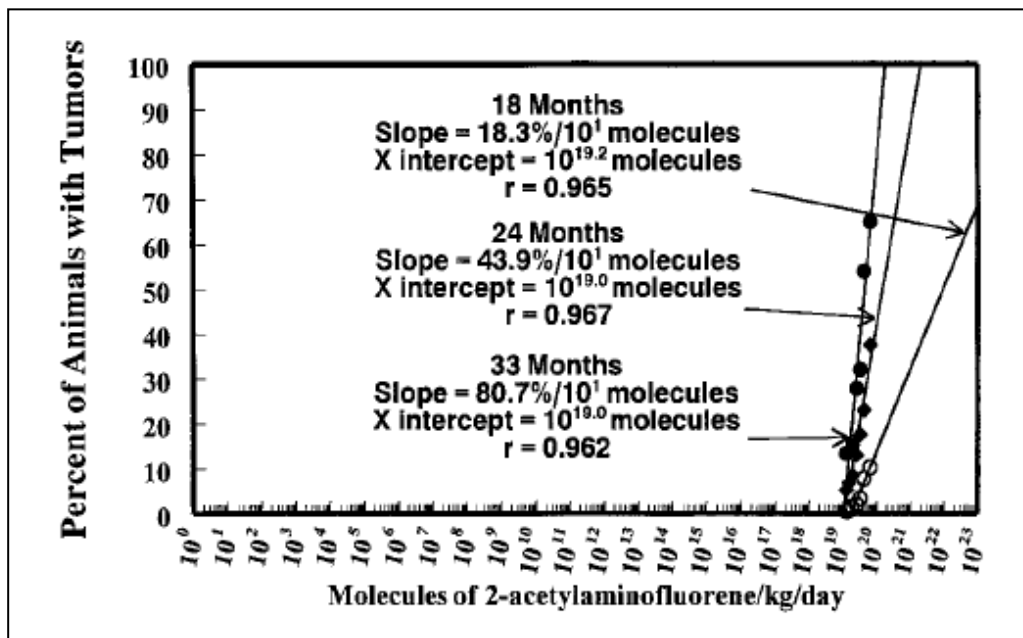
Figure 7: Figure obtained from Waddell [34] shows the percentage of all mice (dead, moribund, and terminated) at 18, 24 and 33 months, with liver neoplasms plotted against molecules of 2-AAF/kg/day. Waddell [34] obtained the data from Table 2 of Farmer et al [43].

The use of the Rozman scale [49] by Waddell in his re-evaluation of the ED01 data received a number of criticisms from members of the scientific community. The analysis was described by Crump & Clewell as highly flawed [50] and Waddell was accused by Anderson et al [51] of drawing unsubstantiated conclusions about the shape of the dose response curve. Some of the main criticisms of the approach employed by Waddell [34] included:

- The inability of the extrapolation method to include data from the experiment control group. Lutz [52] cited this as a major shortcoming of the approach as it was considered that any discussions of dose-response relationships and extrapolation should include the background process.

- Andersen et al [51] criticised the approach adopted by Waddell for not taking into account the biological mechanisms involved in the toxic action of 2-AAF when choosing the metric scale for assessing the dose response. It was their opinion that it is the mode of action that determines the preferred metric for assessing target tissue toxicity. The use of number of molecules on the dose scale was not considered to be appropriate for every situation. It was the opinion of Andersen et al [51] that the chemical potential of the administered compound should not be used as the universal representation of tissue dose in toxicological studies.

- The proposal of the use of a linear model relating log dose (expressed as molecules/kg/day) as a solution to the historical uncertainties associated with the controversy over whether or not a threshold existed for animal carcinogenicity data, was considered to be over simplistic. In particular Haseman [53] criticised Waddell's inability to include all data points especially in the low dose region. The use of only 2 doses by Waddell [35] in his threshold extrapolation was criticised and this was considered to force the existence of a threshold. Haseman [54] asserted that it was not possible to carry out meaningful threshold extrapolation on the basis of just 2 doses particularly when the lowest dose produces a response well above the control levels.

In his response, Waddell [55]rejected as unfounded the criticisms to his analyses of these data sets.

3.2   The Rat Mega study

A second mega study was conducted by Peto et al [26], employing almost 5000 rats to investigate the dose response relationship using two different carcinogens; *N*-nitrosodiethylamine (NDEA) and N-nitrosodimethylamine (NDMA). The experiment was set up to examine the effects of NDEA on the oesophagus and hepatocytes and the effects of NDMA on bile ducts and hepatocytes. The authors administered NDEA and NDMA in drinking water to inbred Colworth rats starting from 6 weeks and continuing throughout life. There were 16 different treatment groups with NDEA and NDMA doses ranging from 0 to 16.896 ppm.

The authors presented the dose response relationships for the different datasets using Weibull distributions intended to provide a simplification of the overall results. An appreciable background incidence of cancer was reported for the liver (liver cancer was observed in 29 of the control animals). No incidences of oesophageal cancer were observed in the control animals or animals treated with NDMA. In their investigation of the low dose effects of the nitrosamines the authors tested 5 doses between 0 and 1 ppm (0, 0.033, 0.066, 0.132, 0.264 and 0.528). More detailed information on the use of the Weibull dose response model was provided by the authors in a separate publication [56].

To investigate the dose response relationship at such low doses, the authors pooled the data for males and females for both NDEA and NDMA (figure 8). The study investigators noted that although it would have been preferable to model the data separately for both sex and the different nitrosamine compounds, even the pooled dataset showed that only 20 incidences of liver tumours were observed at the lowest dose levels [26]. It was noted that separate evaluation of the individual sexes and compounds would have resulted in unreliable characterisation of the shapes of the dose response curves.

The modelled dataset showed an apparent increase in liver tumour incidence after 0.3 ppm. Among the four lowest doses (0, 0.033, 0.066 and 0.132 ppm), a positive trend was observed, although statistical significance (1 $P = 1.7$ %) was only just attained. The latter prompted the

authors to conclude that a linear relationship for low dose response was observed in the liver with no indication of a threshold.

The authors observed that at low doses, incidences of liver (not oesophageal) tumours induced by the treatment were proportional to the dose rate. The relationship below 1 ppm for liver tumours was described as being linear. This was considered to be unsurprising given that an appreciable background incidence of liver tumours was observed among the control groups.

Peto et al [26] acknowledged that while the pooling of such data may be inappropriate, it was justified on the basis that:

- Examination of the pooled dose response data enabled observations of patterns that might have otherwise been drowned in random error in each of the separate dose responses [26].
- Modelling of the specific dose response relationships would have resulted in such few animals being affected that it would have been impossible to reliably characterise the shapes of the dose response relationships below a dose of 1 ppm [26].

It was noted that the pooled data provided some evidence in support a linear relationship for the incidence of liver tumours with dose. The authors reported that the results for liver tumours suggested linearity in the dose range below 1 ppm (0.1-1 ppm). The authors concluded that there was approximate proportionality of excess risk to the applied dose rate and suggested that a dose of 1 ppm will cause about 25% of rats to develop liver tumours, similarly a dose of 0.1 ppm would cause about 2.5% of the rats to develop tumours. Although it could not be observed directly from the experimental data, the authors assumed that a dose of 0.01 ppm could result in a liver tumour incidence of 0.25% and so on.

Figure 8: source- Peto et al [26].

Fig. 13. Pooled dose-response relationships in the low dose range. The "pool" is of both sexes, and of both NDEA and NDMA. ×, esophageal lesions; ●, liver lesions. *Bars*, ranges of rates 1 SD above and below observed rates. Each treated group (but not the controls) began with 240 animals (60 male NDEA, 60 female NDEA, 60 male NDMA, 60 female NDMA), of which 192 were not scheduled for sacrifice at 12 or 18 months. The control group was double this size, so the control numbers are halved and plotted twice.
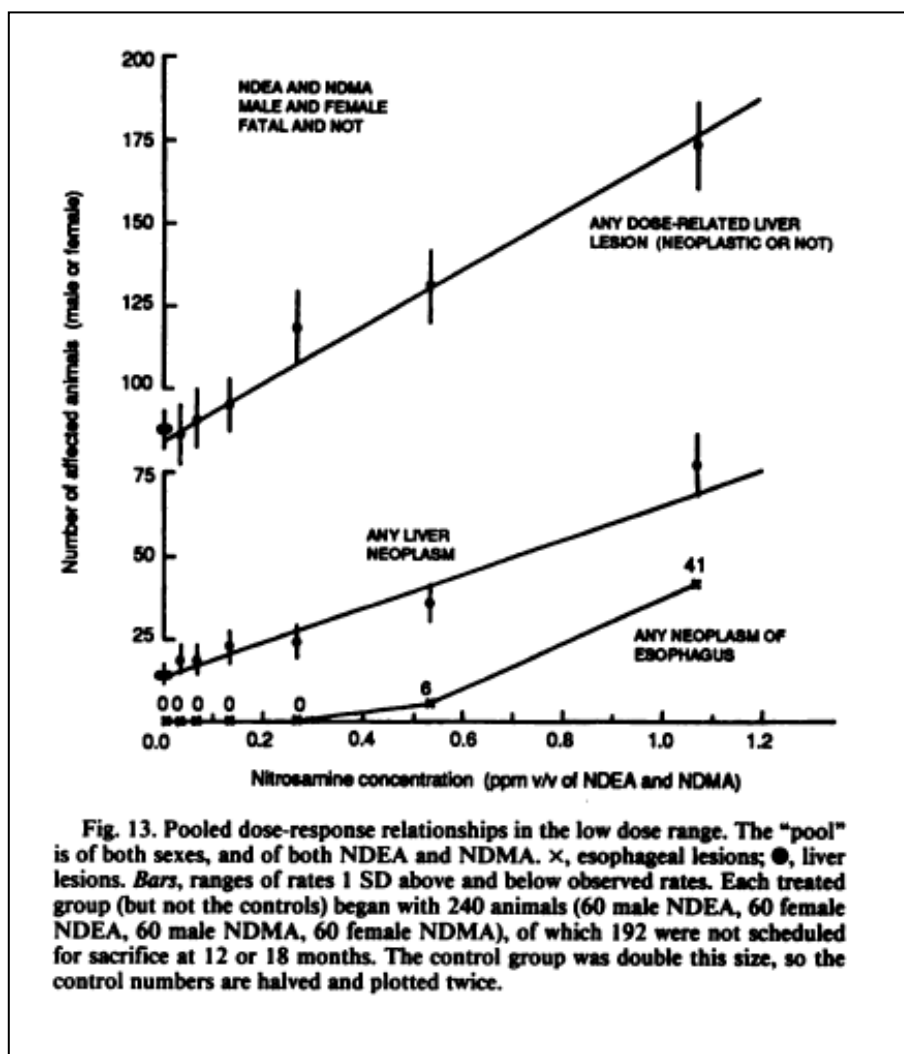
Figure 9: Observed and expected (based on dose-response model used) numbers of animals with lymphatic or hematopoietic neoplasms for male and female rats exposed to NDEA and NDMA (only the three lowest doses administered are shown here).

Source: Peto et al [26]

Treatment group	Concen-tration (ppm)	Control and NDEA, males		Control and NDEA, females		Control and NDMA, males		Control and NDMA, females	
		Expected	Observed[a]	Expected	Observed[a]	Expected	Observed[a]	Expected	Observed[a]
1	0	20.11	22 (17M, 5m)	10.25	10 (9F, f)	14.33	22 (17M, 5m)	10.59	10 (3F, 6F[b], f[b])
2	0.033	4.69	5 (3M, 2m)	2.47	1 (F)	4.14	1 (M)	3.54	1 (f[b])
3	0.066	5.72	5 (3M, 2m)	2.38	3 (2F, f)	3.65	6 (5M, m)	2.94	2 (F, F[b])
4	0.132	6.15	5 (2M, 3m)	1.87	5 (5F)	3.78	1 (M)	2.27	2 (F, F[b])

It is noteworthy however that the authors concluded that the information obtained does not provide reliable information about the effects of nitrosamine exposure in humans. It was concluded that for oesophageal cancer, no predictions could be made about the shape of the dose response curve as the spontaneous rate was immeasurably small for this endpoint. At the low doses investigated, the authors highlighted that the incidence of oesophageal cancer did not appear to be proportional to the doses tested.

The authors speculated however that after taking into account the appreciable background liver cancer rates in the rats (8%), the linear relationship observed in the dose range of 0.1-1 ppm made it probable that a similar relationship (although impossible to test) would be observed at even lower doses for Colworth rats, if not also for humans [26].

Peto et al concluded that the liver cancer data were not sufficiently reliable to determine the shape of the dose response curve.

Interestingly, the data produced by Peto et al [26] were re-examined by Wadell [35] using the logarithmic scale proposed by Rozman et al [49] in which the dose was represented in molecules/kg/day while the response was represented in a linear scale. It is worth noting however that the latter approach was heavily criticised in a subsequent reanalysis of the ED01 (mega mouse) data, discussed above [51-54]. Although Waddell [35] concluded that a threshold existed for oesophageal cancer ($10^{17.1}$ molecules of NDEA/kg/day), the approach was challenged on the basis of the data being compressed and it was argued that the dose response curve obtained was a visual artefact [50].

3.3   The Trout mega study

Bailey et al [38] put together what was termed the ED001 study. The study investigators used an aquatic carcinogenesis model intended to investigate the dose-tumour response data below the 5 or 1% level of cancer incidence. The authors used the rainbow trout (*Oncorhyncucs mykiss)* for their study which were reported to have historic background liver and stomach cancer rates around 0.1% near the facilities used by the study investigators. The authors reported that the costs of running the aquatic models were a lot lower than the costs required for the upkeep of rodent models.

On the basis of well established pathologies and protocols for carcinogenesis experimentation, the authors of the study sought to investigate whether the rainbow trout model could provide robust dose response data close to or beyond its historic background rate of 0.1% liver cancer rate.

Bailey et al [38] highlighted that one of the difficulties in mega studies that have previously been carried out is the lack of statistical power to calculate the tumour incidence level below 1%. In particular reference was made to the mega rat study carried out by Peto et al [26]

which utilised 4080 animals to study the effects of NDMA and NDEA in Colworth rats. It was noted that the study did not have sufficient statistical power to investigate the carcinogenic effects in liver below 5% cancer incidence, owing to the fact that this was the background cancer incidence for the particular rat strain used in the study.

The authors of the study used biomarkers of cancer risk, such as initial target organ carcinogen-DNA adduct levels, which are intended to provide an indication of the eventual tumour outcome [38]. They therefore incorporated dibenzo[*a,l*]pyrene (DBP)-DNA adduct measurements into their study to investigate low incidence biomarker correlations [38].

The tumour incidence data was fit using nine different models (the Ryzin Rai model and eight dichotomous models in the EPA Benchmark dose software: Gamma, Gompertz, Wiebull, multistage, logistic, log-logistic, probit and log-probit models).

The authors commented on the use of different scales noting particularly the linear scaling approach used by Waddell [34] which they highlighted was challenged on the basis of asymmetric data compression artefact [34, 38]. In acknowledgement of the problems associated with linear-linear scales, most notably the compression of data at the lower end of curve, the authors opted to use a log-log plot to analyse the total trout cancer incidence data. The log-log model used by the authors was reported to avoid data compression and provide a more sensitive scaling method for examining the EPA default assumption which is considered to be conservative[3].

The authors investigated the relationship between the three best fitting models ($p > 0.8$ for each model; Ryzin-Rai (3 df), linear logit quadratic (3 df) and probit linear (4 df)) for the data obtained for liver and stomach cancers (modelled on a log-log scale) and compared it to the default LED_{10} (linear extrapolation derived solely from higher doses of the cancer data). The fitted models incorporated both the background response and dose response to the incidence data.

The Ryzin-Rai model was observed to be the best fitting model with a calculated slope parameter of 2.278 with an asymptotic standard error (SE) of 0.356 which was considered by the authors to demonstrate sublinearity and direct evidence against the claim of direct proportionality at low dose. The authors were unable to obtain single slope parameters for both linear logit quadratic and linear probit to describe the low dose behaviour but observed that the slopes were steeper for both these models than that observed for the Ryzin-Rai model. The authors therefore concluded that all three models fit against the liver data were incompatible with the default EPA model which required that the response observed remain proportional to the dose administered at low incidence.

---

[3] The EPA linear modelling approach is considered to be conservative as it assumes that the exposure related response is directly proportional to the carcinogen exposure dose.

Liver cancer:

A 5 log-order magnitude of extrapolation was required from the default LED_{10} curve from the high carcinogen doses down to the dose expected to result in a one in a million risk of cancer (ED_{10}^{-6} equivalent to 0.126 ppb DBP). In comparison, less than a 3-log order magnitude of extrapolation was required from the 0.02% incidence which could be detected in this study, to obtain the dose that corresponded to the ED_{10}^{-6} level of cancer risk from the $LED_{0.1}$ curve. This was compared to the virtually safe dose estimates of 66 and 186 ppb obtained from the Ryzin-Rai and linear probit models respectively. The differences obtained from the different models equated to almost 500 – 1500-fold greater i.e. less hazardous, than the dose predicted to be the virtually safe dose from the default LED_{10} curve. The authors concluded that the analysis provided some evidence for the degree of conservatism in the use of the default linear extrapolation.

The attempts made by the authors to use DNA adducts as biomarkers of tumour dose response proved to be somewhat unreliable. The authors observed that the dose response slopes for the initial carcinogen-DNA adducts and the eventual hepatic tumour response to DBP were markedly different. It was observed that the total genomic DBP-DNA adducts increased as a power function of the DBP dose (rather than linearly). This dose response relationship was modelled as a straight line with slope 1.31 on a log-log scale (SE, 0.061; 95% CI 1.19, 1.44; $p < 0.007$). This compared with the slope obtained for the DBP-tumour incidence line of 2.28 (Ryzin-rai; SE, 0.356 $p < 0.0001$). It was therefore not possible to reliably predict the shape of the resulting DBP dose-tumour incidence curve.

For risk illustrative purposes, the authors nevertheless used the data obtained from the calculated slope for the DNA-adducts to predict the shape of a theoretical dose response curve. The curve was normalised at the 10.1ppm dose point and the shape of the theoretical tumour dose response curve below 10.1 ppm was predicted. By extrapolating to the level of ED_{10}^{-6} the authors derived a VSD of 2.5 ppb DBP. This value was 20-fold greater than the VSD obtained from using the LED_{10} approach. Although the level of conservatism of the linear extrapolation (LED_{10}) method was highlighted yet again when compared to the data obtained from the DNA-adducts, the level of conservatism was significantly lower (at 20-fold) than the levels of conservatism estimated from the use of actual tumour data (500-1500-fold from both liver and stomach cancers). However, this does assume direct dose concordance between adduct formation and tumour induction, which is highly unlikely.

The authors concluded that DNA-adducts were not reliable biomarkers for ultralow dose risk evaluation owing to the fact that the level of conservatism for the use of the LED_{10} was not adequately captured by the DNA-adduct data. Although less conservative than using the linear extrapolation approach, the use of DNA-adduct data for the calculation of the VSD would also have led to a conservative estimate of the cancer risk, in comparison to the calculation of the VSD estimated on the basis of extrapolations from models fitted around tumour incidence data in this study.

Stomach cancer:

The authors described a hormetic or J-shaped dose response curve for stomach cancer. Caution was however expressed over the interpretation of the shape of the curve as being J-shaped owing to the fact that the single data point which determined this shape was observed at the lowest DBP dose which was noted not to be statistically different ($P < 0.5$) from the responses observed at 0 or at 1.27 ppm [38]. The best fitting models for the stomach cancer data were also found to be Ryzin-Rai, linear probit and linear logit. The authors also fitted the LED_{10} curve to the data for comparison. The authors reported greater variations in the four lowest doses of DBP in both the data replicates and the model fits when compared to observations in the liver cancer data.

The authors attributed the differences observed between the two endpoints to the greater optimisation of data carried out for the liver and also to the fact that the liver had two thirds of the background incidence of cancer and a less steep DBP response compared to the response for stomach cancer. The authors cited this as one of the limitations of this dataset highlighting that the model fits for the stomach data were determined primarily by the high dose, as responses at the lower doses were considered to be imprecise. Regardless of the limitations the authors were able to conclude that both the liver and stomach data were not compatible with the EPA default assumptions [38].

Table 4: Doses extrapolated from the use of the different models

Source: Bailey et al [38]

Models fitted	Liver cancer	Stomach cancer
	VSD -ED_{10}^{-6} (ppb)	VSD -ED_{10}^{-6} (ppb)
LED_{10}	0.126	0.326
Ryzin-Rai	66	460
Logit type model	158	120
Linear probit (log dose)	186	1813

The authors observed that slopes estimated at low-dose for stomach cancer were above 1 (1.97; SE, 0.071 and 2.743; SE, 0.454 obtained from the logistic linear model and Ryzin-Rai model respectively) which was an indication of sublinearity and argued against direct

proportionality ($P < 0.0001$ likelihood ratio test that parameter $= 1.0$). The authors observed a greater range of virtually safe dose estimates for stomach cancer compared to liver. This difference was put down to the greater dispersion in the low-dose region of the stomach data.

Similar to observations for the liver data, the virtually safe dose (VSD) estimates obtained from the stomach data showed that they were in some cases over 5000 times greater than the VSD estimated using the linear default model.

One of the main advantages cited by the authors of the study was the ability to reduce the level of extrapolation necessary from the low dose exposures. It was noted that the default linear extrapolation methods require extrapolation by 5 orders of magnitude below the modelled dataset.

Relevance of biological model

Although the authors highlighted the use of a non-mammalian model as a potential limitation in their study, particularly for the calculation of the $ED_{10}{}^{-6}$, they were keen to note that their intention was not to extrapolate the data from trout to either humans or rodents. Rather, the goal of the study was to provide experimental data that explored the incidences of cancer at ultralow exposures. In addition, the data were used for the assessment of dose linearity for tumour response and to investigate the extent to which DNA-adducts could be used as biomarkers of tumour response at low dose.

3.4   *Other Studies of Genotoxic Carcinogens*

In addition to the compounds evaluated in the mega studies, four other genotoxic compounds have been investigated in sufficient detail to contribute meaningfully to this review of the nature of the dose-response curve at low exposure levels. These studies are briefly discussed here.

2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP)

Shirai et al [30] carried out tests of medium and long term duration to investigate the effects of PhIP in rats on mammary carcinogenicity. Female F344 rats were administered doses of 25, 100 and 200 ppm of PhIP in the diet for 104 weeks or 400 ppm for up to 52 weeks. In addition to lesions in the colon and mammary glands, the authors also observed lesions in the skin and subcutaneous tissues.

The authors carried out separate studies in Sprague-Dawley rats administering doses of 25, 100 and 200ppm for up to 48 weeks. They reported a clear dose-response relationship in both the latent period of tumour appearance and the final tumour incidence. The authors chose to

investigate the preneoplastic effects in male F344 rats in a medium term bioassay which involved the use of DEN (200 mg/kg. i.p.) as an initiation agent. The formation of GST-P positive foci and DNA adduct formation were taken as the endpoints in this assay. The formation of adducts was dose dependent and the authors showed that the formation of DNA adducts did not correlate with tumour development.

Although the authors did not provide a thorough dose response analysis, they calculated the VSD at a risk level of 10^{-6} by using the Weibull model (0.023-0.52 ppm in diet). The dose that equated to this risk level was estimated to be 0.4 to 16µg/day (7.5 x 10^{-5} to 10^{-3} ppm) in humans.

Vinyl Chloride (VC)

Zapponi et al. [25] investigated the reproducibility of risk estimates among different rodent species and strains following low dose extrapolations. The authors evaluated seven different studies and noted that the experimental designs were comparable in all the studies. The studies involved two different rat strains and two different mouse strains which enabled the authors to carry out interspecies and inter-strain comparisons. The doses investigated in the various studies ranged from 50 to 2500ppm. They carried out the dose response analysis using the Weibull model and described the dose response as having a characteristic downward curve (supralinear) in the range of experimental doses. The authors concluded that the downward trend observed for the curve was "not consistent with the hypothesis of a threshold".

3.5 Dose response investigations based on preneoplastic endpoints

Five of the 17 key studies identified sought to investigate the dose response relationships of genotoxic carcinogens using preneoplastic endpoints. These studies are discussed in more detail below.

Aflatoxin B1(AFB₁)

Choy [29] reviewed dose response studies on the induction of DNA adducts by aflatoxin B1. The author analysed both intraperitoneal (IP) injection studies and ingestion studies. Although the author examined 13[4] different single dose IP studies that applied doses that ranged from 2.5x10^{-7}-2 mg/kg, generating a composite graph for the data, statistical analysis of dose-response data was carried out on the basis of just a single study, that of Appleton et al

---

[4] It should be noted that none of the 13 studies reviewed in this paper were picked up by the search strategy developed for this review.

[57] in which the authors investigated the effects of AFB_1 on male rats at 6 dose levels that ranged from 10 to 10^3 ng/kg. Choy felt unable to analyse the combined data because of its heterogeneity. Statistical analysis of the data from Appleton et al showed that the response relationship for DNA adducts was linear to as low as 1 ng/kg bw, the lowest dose administered. Choy also reviewed 6 ingestion studies in rats that were conducted in the low dose region (1 to 2.5 x 10^5 ng/kg). Choy [29] constructed a composite dose response curve on the basis of the 3 studies in which a single dose was administered. The dose response relationship was linear to as low as 1 n/kg bw, the lowest dose tested.

Diethylnitorosamine (DEN)

Williams et al [28] investigated neoplastic changes that were induced in the livers of 108 male F344 rat livers administered diethylnitorosamine (DEN). Doses of 1, 2, or 4 mmol/kg body weight of DEN were administered in the diet over a 34 week period. The authors monitored the levels of altered hepatic foci and different histochemical markers in the liver. The latter included the levels of glutamine synthetase (GS^+) which is thought to be a cell-lineage marker for foci from the small perivenous hepatocyte population that express this phenotype (Gebhardt et al [58]; Williams et al [28]). Altered hepatic foci were selected by the authors as a parameter for the dose response analysis as the numbers of induced foci are thought to be predictive of the development of neoplasms with the same phenotype (Williams, [59]; Williams et al., [28]). The authors employed the use of phenobarbital (PB) as a promoter to enhance the development of the foci to neoplasms. The authors also monitored the levels of γ-glutamyltranspeptidase (GGT) and iron storage (IS) within the foci observed, as altered foci are thought to display abnormalities in these two phenotypic characteristics.

The authors did not report carrying out any formal modelling of the dose response data; instead they used the magnitude of multiplicity of foci at the different dose levels as an indication of the dose response relationship for the endpoints of interest. They reported a non-linear relationship for the development of altered hepatic foci. The authors also reported a non-linear dose response relationship for neoplasms as no neoplasms were observed in the low dose groups studied. As expected, the authors observed that all DEN induced foci could be characterised by changes in IS^- and GGT^+, the multiplicity of which were observed to be dose related. GS^+ foci were observed to be present only in the high dose group.

2-Acetylaminofluorene (2-AAF)

Poirier et al [27]- Mouse

Poirier et al [27] carried out DNA adduct analysis using a similar protocol to that used in the ED01 study. The objective of the study investigators was to compare their DNA adducts data with the tumour data from the ED01 study, in an attempt to determine the number of "adduct related" events that are necessary for the induction of liver and bladder tumours. The levels of

DNA adducts in mouse liver and bladder were quantified after 28 days of AAF feeding at 9 different dose levels (0, 5, 10, 15, 30, 45, 60, 75, or 150 mg/AAF kg diet) then compared to the ED01 reported incidences of neoplastic (carcinomas) and preneoplastic (adenomas for liver and hyperplasia for bladder) lesions after 24 months of AAF feeding.

The study investigators subjected the data to Weibull analysis and found that there was a linear correlation between the administered dose of AAF (in the ED01 study) and the concentration of adducts (dG-C8-AF) in both the livers and bladders of treated mice. However they noted that there was a clear difference in the dose response relationship for tumour induction and adduct formation at the different tumour sites, where a linear response was observed for liver tumours and a non-linear response was observed for bladder tumours. The authors interpreted the observed linearity between tumours and adducts to be the result of the liver requiring only one AAF-related event to induce cancer in this tissue, whereas the non-linearity in adduct-tumour relationship observed for the bladder was interpreted likely to be the result of the need for multiple carcinogen-related events (e.g. cell proliferation) to induce cancer in this tissue.

The authors noted that their study clearly demonstrated that the relationship between DNA adduct concentration and tumourigenesis was tissue specific. The authors added that the differences observed between the two endpoints provided further evidence that DNA adducts are necessary but not sufficient for tumourigenesis.

Williams et al [32]- Rat

Williams et al., [32] administered 2-AAF to F344 rats to investigate the existence of potential thresholds for carcinogenesis in the liver. To ensure precise dose administration of 2-AAF, the authors administered the doses via intragastric instillation (IG). The authors administered doses between 0 - 448 mg/kg but the doses were expressed as a cumulative total. The authors tested for the effects of 2-AAF on formation of DNA adducts, arylsulfotransferase, glutamine-synthetase (GS) zone, replicating fraction (RF) and hepatocellular altered foci (HAF).

The dose response relationships were not displayed in the published paper, the authors chose instead to describe the dose response relationships in terms of the induced changes at the different cumulative doses of exposure. The authors observed that the enzyme aryl sulfotransferase in the liver was not inhibited at the cumulative low dose (CLD) of 112 mg/kg/day but was inhibited at the cumulative mid dose (CMD) and cumulative high dose (CHD: 448 mg/kg/day).

A nonlinear dose response relationship for the formation of DNA adducts was described and this parameter was the most sensitive of the endpoints assessed. No DNA adducts were formed at the CLD so a NOEL of 0.094 mg/kg was identified for the formation of DNA adducts and the threshold was said to be 28.2 mg/kg.

The level of hepatocellular proliferation was measured through the investigation of the RF. The authors observed a marked compensatory increase at the CHD in the expression of RF at 8 weeks but not at 4 weeks. HAF expressing GST-P were induced at the CHD by 4 weeks but not at the CLD and a NOEL of 28.2 mg/kg was also identified for this endpoint. The dose response relationship for the effects on HAF were described as being supralinear at the CHD studied (282.2 mg/kg).

The authors concluded that their study was an important illustration of the differences of preneoplastic effects observed at high and low doses of exposure. They suggested that linear extrapolation from effects at high dose to postulated effects at low doses could lead to the over-estimation of the level of effects or risks. The authors therefore concluded that such extrapolations must be supported by mechanistic data.

Methylmethane sulfonate (MMS), Methylnitrosourea (MNU), Ethylmethane sulfonate (EMS) and Ehtylnitrosourea (ENU)

Doak et al [36] investigated the biological significance of low dose exposures to four alkylating agents MMS, MNU, EMS and ENU by quantifying the levels of chromosomal damage and point mutations. The *in vitro* study employed the use of human lymphoblastoid cells and the parameters (chromosomal aberration induction and point mutation frequency) were quantified using the cytokinesis blocked micronucleus assay (CBMN) and hypoxanthine phosphoribosyltransferase (HPRT) forward mutation assays. The aim of the authors was to determine whether or not a NOEL could be identified from the low concentrations to which the cells were exposed. For the CBMN assay the authors exposed the cells to 0-0.25 µg/mL (MMS, EMS and ENU) and 0-0.08 µg/mL (MNU). For the HPRT assay the authors exposed the cells to 0-2.0 µg/mL of all four carcinogens.

To determine the significance of the induced changes observed in the treated cells compared to non-treated cells, the authors applied a one-way ANOVA test followed by a Dunnett's post hoc test. Statistical modelling was by root transformation of the data after the control values had been subtracted to standardize the values. The values were then plotted on a log scale to determine the NOEL or LOEL concentrations.

The authors observed a linear dose response relationship for the induction of chromosomal damage and point mutations by MNU and ENU, although it was not stated whether the overall trend observed was statistically significant. Effects were observed at all concentrations, although the first concentration that produced a statistically significant response was 0.15 and 0.5 µg/mL respectively ($p < 0.05$) for the induction chromosomal damage and 0.0075µg/mL and 0.4 µg/mL (MNU and ENU respectively) for the induction of point mutations. The authors concluded that it was not possible to identify a NOEL for either MNU or ENU.

Although it was noted that some of the induced genetic changes at the lower concentrations were not statistically significant, the authors acknowledged that the potential for these chemicals to induce effects at these levels could not be dismissed. In contrast exposures to EMS and MMS were shown to result in nonlinear dose response relationships. The authors reported LOELs of 0.85 and 1.4µg/mL for MMS and EMS respectively for the induction of chromosomal damage and LOELs of 1.25 and 1.4µg/mL for the induction of point mutations for MMS and EMS respectively. Importantly it was noted that the LOELs were at doses that did not significantly decrease cell viability. The authors therefore concluded that the sharp increases in chromosomal damage and mutation frequencies observed were not induced via a cytotoxicity related mechanism. Clear NOELs were identified for both EMS and MMS.

Gocke and Muller [37] subsequently carried out a study to determine if the effects described above could be observed in vivo. Mice were treated with EMS at doses of 0, 1.25, 2.25, 5, 20, 80, 140, 200 and 260 mg/kg/day or ENU at doses 0, 1.11, 4.45, and 17.8 mg/kg/day. The authors investigated the mutagenic and clastogenic activities induced by the two chemicals and this was carried out through the use of the micronucleus test (MNT), after treatment for 7 days and a gene mutation test (Muta™Mouse), after treatment for 28 days.

The authors carried out a four step statistical assessment of the genotoxicity data which included:

- Comparison of control groups

- Rejection of linear dose-response relationship (entire dose range).

- Acceptance of linear dose-response relationship below the NOEL.

- Application of threshold software developed by Lutz and Lutz [60] to calculate the threshold doses including confidence limits.

The dose response curve was then plotted using linear regression. The results obtained from this in vivo study were consistent with the observations reported by Doak et al [36]in their in vitro study. Gocke and Muller [37] observed a clear threshold for EMS in the dose response for the induction of micronuclei in the bone marrow of the mice tested, with a NOEL of 80 mg/kg bw per day. Cytotoxic effects were not observed until the highest dose of 260 mg/kg/day. Levels of ethylvaline adducts in blood increased continuously with dose, but only once they reached 100 nmol/g globin was there any increase in chromosomal damage.

In accordance with the observations of Doak et al [36], with ENU there was an increase in MN induction (although not statistically significant) at the lowest dose tested (1.11 mg/kg/day) and the dose response curve for this carcinogen was linear, with no threshold. Again, this was not due to cytotoxicity.

The Muta™Mouse test showed that ENU induced point mutations at the lowest dose tested (1.39 mg/kg/day) in bone marrow, liver and gastrointestinal tract, the three tissues studied. The dose-response curve in all three tissues was essentially linear, although it should be noted that there were only three dose groups in this part of the study. In contrast, EMS induced no gene mutations in the GI tract or bone marrow at doses up to 25 mg/kg, or up to 50 mg/kg/day in the liver, above which there was a modest increase in mutation frequency, respectively. The authors calculated that mouse liver could tolerate up to 380,000 DNA adducts per day without induction of mutation, presumably because of DNA repair.

IV. Discussion

The main aim of this study was to carry out a systematic review to critically evaluate available evidence concerning the empirical form of dose response relationships at low dose exposure to genotoxic carcinogens in animals, to help in interpreting the level of concern for the margin of exposure. The first step of the review was literature retrieval. Inclusion criteria included: published papers, known genotoxic carcinogen, animal study (although some key in vitro studies were also included), information on dose-response relationship. Due to the inconsistent use of key words within PubMed in the literature of concern, it was not possible to conduct an efficient yet comprehensive systematic search relying on the use of search strings made up of specific key words to identify relevant publications. Additional relevant papers were therefore identified using citation searching, ad hoc searches in PubMed and also through the solicitation of suggestions from experts within the field of low dose cancer risk assessment (identified from the participants of the elicitation workshop carried as part of objective 03/01 of this project).

The strategy finally adopted yielded a large number of papers (several thousand), which were then screened for relevance. This led to the identification of 17 papers which were reviewed in detail. Ten chemicals were investigated amongst these 17 papers. The chemicals fell into the IARC categories of group 1 and 2 human carcinogens and included: aflatoxin B1 (AB), 2-acetylaminofluorene (2-AAF), diethylnitrosamine (DEN), ethyl methanesulfonate (EMS), ethylnitrosourea (ENU), N-nitrosodiethylamine (NDEA), nitrosodimethylamine (NDMA), dibenzo[a,l]pyrene (DBP), 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) and vinyl chloride monomer.

In addition to the focus on studies that evaluated the low dose responses to genotoxic carcinogens in animals, one of the objectives for the review was also to consider the types of toxicological endpoints that were being evaluated by the study investigators and the types of statistical models employed for cancer risk estimations/dose response extrapolations. As such, the studies were divided into 3 different categories:

- Papers that investigated tumorigenic responses

- Papers that investigated preneoplastic relationships

- Papers that investigated genotoxic responses that may not be directly causal in cancer[5]

Preference was given to the following studies:

- Studies that provided experimental evidence for the nature of the dose-response at low doses.

---

[5] This classification is based on evidence from the literature and the current consensus on the use of some of these endpoints as being directly linked to the carcinogenic effects of genotoxic carcinogens.

- Studies that presented new data, and were not a repeat of previously published information (although several important revaluations of previous studies were included).
- Studies that investigated endpoints that is relevant to genotoxic carcinogenicity.

There were only three major studies on the tumourigenic effects of genotoxic carcinogens, over a wide dose range, that were sufficiently well powered to detect an increase in incidence below that normally observable in a rodent cancer study (typically 10%). These were the mouse mega studies published by the National Centre for Toxicology (NCTR) [1, 43, 44], the rat mega study [26, 56] and the trout mega study [38]. The results published by the NCTR were the largest ever conducted rodent study. Also known as the ED01 study, it was designed to estimate with precision the effective dose that produced a one percent increase in tumour incidence above the spontaneous background level [40]. One of the benefits of the study was that it led to the generation of a substantial amount of data that would later be made available to the scientific community. Owing to the logistics and financial costs of conducting such a large chronic rodent study, the dataset produced from the ED01 data has proved invaluable to members of the scientific community as it is unlikely that such a study will ever be repeated.

The ED01 study conducted by the NCTR was the first of its kind to investigate the shape of the dose response curve at such low levels. Although the aim of the study was to investigate the dose that produced a 1% tumour incidence detectable above the spontaneous background rate, and not to investigate whether genotoxic carcinogens exhibit thresholds in their effects, the scale of the study meant that it gained a lot of attention from the wider scientific community, not least because of the potential political and regulatory implications of the findings.

Application of statistical models and shapes of dose response curves

For only five of the chemicals was sufficient information generated to adequately describe the shape of the dose response curve. These were 2-AAF, NDEA, NDMA, aflatoxin B1, DMBA and DBP.

2-AAF

2-AAF was the most analysed compound among the papers obtained for the review. This was mainly due to the fact that it was selected as the case compound for the mega study conducted by the NCTR [1] which resulted in the production of a substantial amount of data available within the public domain. Following the original publication by Littlefield et al [1], a number of further publications were produced by various authors [2, 23, 24, 27, 34, 50-52, 54, 61] including one by representatives of the NCTR [62] in response to some of the re-analysis of the ED01 data. Relevant studies on 2-AAF were also published by other authors.

The dose response analyses carried out on 2-AAF were based on tumour responses (liver and bladder) and preneoplastic endpoints that included DNA adducts, cytotoxicity, compensatory hepatocyte proliferation and formation of preneoplastic hepatocellular altered foci (HAF).

The difference in the outcome (Table 5) from the various dose response analyses of data for 2-AAF provides a clear illustration of some of the difficulties inherent in low dose response analysis of genotoxic carcinogens. The four different re-analyses carried out on the ED01 data produced very different results and conclusions.

In their conclusions, the NCTR [1] stated that a clear linear relationship had been observed for liver tumours and that the incidence of bladder tumours dropped off sharply as the dose of 2-AAF was reduced. They also added that the data for bladder neoplasms did not contradict the "no threshold" theory and the liver neoplasms data provided support for it. The conclusions reached by the NCTR came under criticisms from members of the scientific community. The issue was formally raised and discussed by the Society for Toxicology [45, 47].

The subsequent reanalysis of the ED01 data by the SOT and various other authors highlighted the contentious nature of issues raised by the results of the ED01 study. Although the re-analyses were based on the same data, the impact of the choice of model on the results obtained is perhaps best illustrated by the re-analysis carried out by Purchase and Auton [2]. The main point of contention regarding the ED01 study was over the existence of a threshold, particularly for the bladder tumour results. Purchase & Auton [2] showed that the same data plotted on different scales (log/log and log/linear) produced very different conclusions. Although the log/log plot provided no evidence of a threshold for the bladder tumour data, the log/linear plot provided a clear indication of a threshold.

Table 5: Differences in the results obtained from the different dose response analyses of studies with 2-AAF

Compound	Author	Endpoint	Shape described			Threshold claimed?
			Linear	Non linear	Other	
2-AAF	Littlefield (1979)	Tumours (bladder & liver – ED01)	✓			No
	Gaylor (1985)	Tumour (Bladder – ED01)	✓			No
	Waddell (2003)	Tumours (ED01 reanalysis)			✓	yes
	SOT (1981)	Tumours (ED01 reanalysis)		✓		yes
	Purchase & Auton (1995)	Tumours (ED01 reanalysis)	✓			No & yes
	Poirier (1991)	DNA adducts (ED01 reanalysis)	✓			No
	Williams (2000)	DNA adducts		✓		yes
	Williams (2004)	DNA adducts		✓		yes
		HAF, hepatocyte proliferation			✓ (supr alinear)	yes

By far the most disputed [50-54] re-analysis of the ED01 data was that of Waddell [35] in which the author claimed that a plot of the ED01 data using the Rozman scale [49] showed a clear and consistent threshold not only for bladder tumours but also for liver neoplasms. By contrast, although there were criticisms (SOT, 1981) over the statistical methods applied in the original analysis of the ED01 data carried out by Littlefield et al [1], no other author disputed the lack of observed threshold for liver tumours. The application of the Rozman scale [63] by Waddell [34] led to the description of the threshold in terms of molecules/kg/day: $10^{19.5}$ and $10^{19.1}$ molecules/kg/day for bladder and liver neoplasms, respectively.

One of the main criticisms of the ED01 study raised by a working group of the SOT [24] was on the models used for the original analysis. The SOT working group suggested that the Hartley-Sielken model provided a better fit to the ED01 data and concluded in their

reanalysis that the dose response for both liver and bladder, was non-linear following analysis using time-to-tumour techniques. In addition the SOT concluded that their re-analysis of the data confirmed that a threshold existed for bladder tumours at 60ppm in the diet. In a subsequent publication by scientists from NCTR (Kodell et al, [62]), the SOT criticisms of the original data analyses of the ED01 study were rebutted.

The disputed findings of the initial publication by Littlefield et al [1] with regards to the dose response analysis for bladder tumours, in which it was stated that "the total results were consistent with a no threshold concept", led to a revaluation of the dataset by Gaylor et al [23] to clarify their view on the perceived "impression of a threshold" by others. This involved further pathological evaluation of slides from the original study and dose-response analysis using more specific diagnostic criteria for the bladder lesions. The authors reaffirmed their view that there was no threshold in the dose response relationship for bladder tumours. However, they also stated that "the shape of the low dose response curve for bladder carcinomas remains uncertain".

The fact that after over two decades Scientists have continued to re-analyse the ED01 data is testament to both the monumental efforts that were put into the original study and the difficulties inherent in the analysis of the low dose region of the dose response curves for genotoxic carcinogens. More recent efforts in the interpretation of the ED01 study and otherwise on investigations into the effects of low dose exposure have focused on the use of non-tumour, pre-neoplastic parameters, such as DNA adducts.

Three papers were obtained that investigated the dose response relationships of low exposure to 2-AAF on DNA adduct formation. Of most interest was the study carried out by Poirier et al [27] which compared the ED01 data with the formation of DNA adducts in an attempt to determine the number of "adduct related" events that are necessary for the induction of liver and bladder tumours. On Weibull analysis of the data, DNA adduct levels were linearly related to the tumourigenic response in liver but not in bladder of mice fed continuously with 2-AAF. The authors concluded that their study clearly demonstrated that the relationship between DNA adduct levels and tumourigenesis was tissue specific. They noted however that the differences observed between the two endpoints provided further evidence that DNA adducts are necessary but not sufficient for tumourigenesis.

Williams et al [32, 33] published a series of papers that examined the dose response relationships and the notion of thresholds for tumour induction in rats treated with 2-AAF. The doses administered were expressed as cumulative doses. The authors reported a non-linear dose response relationship for DNA adducts at low levels of exposure while a plateau was found at toxic levels of exposure (cumulative dose of 94.1 mg/kg).

Williams [32] did not observe a quantitative relationship between the formation of DNA adducts and the formation of tumours. The relationship was non-linear, with a NOEL of 0.094 mg/kg of cumulative exposure for the formation of DNA adducts in rat liver after exposure to 2-AAF. The dose response curve for liver tumour formation was also non-linear, with a NOAEL of 28.2 mg/kg (cumulative dose). The differences observed at high to low

dose exposures were considered an indication of the different mechanisms in play at the different levels of exposure. As such the authors questioned the validity of the use of linear extrapolation methods highlighting that their results showed that such methods would lead to overestimations of cancer risk.

Low dose response relationships for N-nitroso compounds

Six different publications that investigated dose response relationships for low dose exposures to N-nitroso compounds were identified. The compounds investigated were NDEA, NDMA, DEN, EMS and ENU. The largest study was the rat mega study carried out by Peto et al [26] who administered low doses (0 to 16.896 ppm) of NDEA and NDMA to almost 5000 inbred rats in total. The authors illustrated the dose response relationships for the effects of the two compounds on the incidences of liver and oesophageal cancer using Weibull distributions. The authors reported a linear dose response relationship for liver though it must be noted that this was based on pooled data (for males and females for both NDEA and NDMA). The authors justified the pooling of the data by stating that it necessary to enable characterisation of the shape of the dose response curve, as there was insufficient power when the groups were analysed separately.

Interestingly, Waddell [35] carried out a reanalysis of the rat mega study data using the logarithmic scale proposed by Rozman [49] (in a similar approach to that used in his re-analysis of the ED01 data) and concluded that a threshold of $10^{17.1}$ molecules of NDEA/kg/day existed for the induction of oesophageal cancer. The findings by Waddell [35] were again criticised as being a visual artefact [50].

Waddell et al [64] examined three published datasets on the effects of NDEA in rats [26, 31, 65], to investigate dose-response relationships for tumour formation and the induction of pre-neoplastic changes. Although the different authors used different dosing regimens, including the use of phenobarbital as a promoter [31], the tumour data were treated as directly comparable. For all endpoints, Waddell et al [64] expressed the doses as molecules/kg/day and plotted the data using the Rozman scale [49]. The authors concluded that all of the dose-response relationships exhibited a threshold, with a difference in thresholds for the induction of liver tumours of about 1.5 orders of magnitude. When the data were analysed using the total cumulative dose expressed in terms of molecules/kg, the threshold for tumour induction for the different studies was $10^{20.3}$ molecules/kg, the difference between studies being within the error of the calculation. Waddell concluded that there was also excellent agreement between the two studies for the appearance of GST-p positive foci, with a threshold of $10^{19.5}$ molecules/kg of total cumulative dose and the data for DNA adducts were modelled better using an exponential fit rather than a linear fit. The authors concluded that there was a threshold for the dose-response for adduct formation and that this was lower than that for tumour formation. However, they did not provide a numerical value, because of the model used.

Waddell et al [64] suggested that the use of the cumulative dose metric was in agreement with the calculations of Rozman et al [66] which suggests that it is dose x time that determines the toxicity of a compound. The authors concluded that there was good concordance of data on DNA adduct formation, hepatocellular altered foci and tumour formation when the dose was based on the cumulative dose of NDEA. It was suggested that this provided further evidence of the presence of a definitive threshold for preneoplastic events and tumour formation for genotoxic carcinogens.

Although the conclusion reached by Waddell et al [67] was that there was concordance between the three datasets, one has to question the validity of such a conclusion when such heterogeneous studies are combined. In addition, in his previous publications, [34, 35, 48, 68] Waddell had emphasised the need for doses to be expressed in terms of molecules/kg/day even in spite of criticisms from members of the scientific community about the validity of such an approach. The plotting of the cumulative dose from the three respective studies whereby the doses were expressed in terms of molecules/kg instead of molecules/kg/day, could give the impression of a dataset that is being forced to show thresholds.

DEN was analysed by Williams et al [28, 31, 33] in a series of studies in which rats were administered low doses to examine the effects of the compound on a range of preneoplastic endpoints (DNA adduct formation, hepatocellular altered foci, liver cell proliferation, level of glutamine synthetase (GS+)) and tumours. With the exception of one paper [31] describing the use of least squares regression, no formal statistical modelling for the dose response analysis was described by the authors. Rather, the conclusion of a non-linear dose response shape for the early DEN effects was on the basis of the magnitude of observed effects at the different dose levels. The authors also identified no-effect levels (NEL) or practical thresholds for the initiation of promotable liver tumours [31].

The most recently investigated N-nitroso compounds were EMS, MMS, ENU and MNU. Doak et al. [36] carried out *in vitro* studies to investigate the effects of these compounds in lymphoblastoid cells in which it was clearly shown that the dose response curves for both MMS and EMS exhibited a threshold for the induction of chromosomal damage and gene mutation, with a linear response above the threshold. ENU and MNU in contrast displayed no evidence of a threshold in the response for the same parameters. Gocke & Muller [37] later carried out studies to determine whether such findings could be observed *in vivo* in mice. It was shown that there was a clear threshold in the induction of micronuclei by EMS, above which the dose-response curve was linear. In contrast, ENU showed no evidence of a threshold and the dose-response curve was linear throughout the entire range.

The mega studies examined here provided the best opportunity to thoroughly investigate the shape of the dose response curve for genotoxic carcinogens at low levels of exposure. However, despite their size, the rat and mouse mega studies [26, 39] enabled reliable data to be obtained only to an incidence of 1% (1 in 10^2) above background. This compares with a nominal level of concern associated with the MOE of around 0.01% (1 in 10^5). On a conventional plot of log dose against response prolonged low dose exposure to 2-AAF,

NDEA and NDMA resulted in a linear dose response relationship, although this was less convincing for the induction of bladder tumours by 2-AAF. In contrast, to the rodent studies, the long term study carried out in trout with DBP[38] was powered to detect a tumour response of < 0.1% above background. The dose response curves for the induction of tumours in stomach and liver were non-linear.

Statistical Considerations (mega studies)

The models

Several papers have stated that the probit (normal sigmoid) and logit (log sigmoid) models give the closest fitting result, assessed often by a chi-square statistic. An advantage of such models is the inclusion of the control group (i.e. zero dose) which can provide an adjustment for natural or background levels of tumour incidence. A log transformation, which has been promoted by some authors, should only be employed by evidence that a log-probit relationship exists which can be assessed by the distribution of error terms. Whether such model assumptions have been tested is not always made clear in the studies.

In the description of the three broad classes of models, it is assumed that the experimental data will be limited to tumour incidence over time. However, the often more probative question or issue is time to tumour. Availability of data on time to tumour observation is limited. Various attempts have been made, employing the Weibull distribution and various gamma or normal based multi-stage models. To determine which such data and models would provide a more realistic model of the complex relationship between dose, dose rate, tumour incidence and age at which tumour detection is observed, the experimental data from many studies, including the mega studies is often insufficient. Analyses of even the ED01 study using such methods have not been shown to adequately describe the experimental data.

Assessment of model employment is also difficult for such studies. Many appear to use specially built software with several of the aforementioned models implemented. There are two risks with this: (1) a lack of knowledge or understanding of what has been employed with regard to parameter estimation procedures and assumptions regarding what the data structure is, both of which could impact the results, and (2) the desire to employ every method and use goodness of fit statistics to assess which model best fits. While utilising statistical approaches such as model averaging to inform model selection (as done by Peto et al [26]) allows for incorporation of any prior information though often specified equally a priori, it may falsely suggest the importance of weights is equal or that one is better over the others. Such methods will be influenced by the models considered (many of which may be inadequate) and the quality, sample, and measurement of the data. Assuming that individual behaviour is random (with individual tolerances) means that models with low goodness of fit indices may be true and useful, however, models with high fit indices may be modelling error or random behaviour rather than the true underlying process.

Other considerations which should be considered before employing a mathematical or statistical model include:

i. *Background exposure:* Background exposure is often accounted for or considered by means of the case control. However, these are lab experiments under controlled environments. Humans are often exposed to a mixture of chemicals which may interact with the carcinogen of interest in different ways. Assuming additivity of such carinogens can have a significant impact on low dose risk estimates.

ii. *Population heterogeneity*: Some of the mega studies do not account for population heterogeneity. The ED01 study was based solely on female mice. The Peto et al., study [26], while including both females and males, combined these during analysis to increase statistical power, and involved inbred mice which may have other inherent prevalent inheritable genetic factors that may impact and interact with the carcinogen in a particular way. From gene association studies it has been shown that certain genders and sub-populations are more susceptible to specific diseases.

iii. *Tumour response observation*: Another consideration is the response observed. Some studies have observed multiple 'end' responses. This may take various forms, for example, different tumour endpoints such as liver and bladder or observing DNA adducts as well as tumour outcomes. DNA adducts, whilst they may be a useful biomarker, can also significantly impact risk assessment. If a single target cell or group of cells is observed, this may have different impacts as not all affected cells may result in a tumour, therefore overestimating risk. While providing a biomarker, the timescale from affected cell to tumour is unknown. Current methods do not adequately account for the complexity in dose to neoplasm relationship involving dose to target cell, DNA damage and repair, and other processes leading to cancerous tumour. One of the other limitations of such studies is not only the limiting range of observed dose exposures but the available information for time to tumour. Some rodents were prematurely sacrificed, and the Bailey et al., study [38] did not allow for time to tumour data to be collected as after the month of exposure to carcinogen was conducted, they were sacrificed at a single final time-point eight months later.

iv. *Dose regime*: The emphasis for many of the studies in on the dose level; however the dose regime (including dose rate) may also impact the results. A carcinogen at low doses continuously, or daily or monthly in shorter durations may have significantly different risk estimates as would the way in which the carcinogen was administered, through solids or water. The biomechanistic process in which the body processes the carcinogen may differ due to the ingestion process, and may not be similarly proportional.

Thus, there are several factors which illustrate a much more complex relationship between dose and response or dose, time and response. When considering dose-response analysis and model fitting methods, serious consideration should be given to the fact that different types of tumour outcomes, chemicals and dose levels may result in different relationships. Some

chemicals induce a carcinogenic effect sooner than others, furthermore their chemical interaction with cells may differ from each other. The tumour outcome whether it be liver, bladder, esophogeal, etc. may progress in different ways and at different rates. It is also possible that not only is there a 'threshold' but that different dose levels have different curves. This is can be seen in Figure 1, taken from Littlefield et al., [1], where doses below 100ppm appear to have a concave shape over the range whereas the dose at 150ppm has a more convex shape over the range. There may be a 'threshold' at a dose level where the dose-response curve differs for higher and lower doses.

Thus, when considering dose-response analysis, the aforementioned factors should be accounted for. These are however limited by the available data, study design, data collection method (including sample size). While some of these studies have conducted a study to consider stochastic variation by using replicates; the sample sizes are still inadequate as the replicate data is combined, masking the uncertainty or stochastic variation in estimation. As each of the three mega studies was designed for a different and specific purpose, it is important to consider this when attempting to re-analyse the data and fit new models. The data may not be adequate for the purposes being considered. It should be noted that the data are often insufficient to distinguish between models and to extrapolate from mid-range to low doses and from animals to humans, and to consider time to tumour.

Dose-response relationship or curve analysis in these studies as presented is limited due to two main factors: (1) there are too few data points, and these do not extend across the range of interest, and (2) the data collected are not often presented in their raw and simple form. Papers often present the fitted models without presenting the data, hence making it difficult to assess the shape of the curve even on the few points. The focus is on fitting the model (to these limited data points) and extrapolating, and presenting goodness of fit statistics. The emphasis should be on dose-response relationship rather than goodness of fit statistics which may be misleading and can be affected by such factors as measurement problems (e.g. too few data points), model misspecification and unaccounted for stochastic behaviour. Thus, the analyses of the data and use of different models as presented in many of the papers often impedes or masks the 'true' dose-response curve relationship. Hogan[19] remarks that it is difficult to decide whether these models are really reflecting some underlying biological mechanism or merely acting as curve fitting devices.

The shape of the dose-response relationship depends on the carcinogen, the kind of response (tumour), dose-rate, time scale of tumour response observations and the experimental subject (animal or human). The statistical form of the dose-response model selected is an important consideration, as different models can produce very different estimates of risk outside of the experimental range of exposure levels. Careful consideration should be taken to assess whether the appropriate scale is being employed (e.g. log, log-linear, etc.) and what may be masked or misused by assuming certain measurement scales and models. Since the models do not adequately model the biomechanistic process, risks are assessed by extrapolating to low doses. Future approaches should carefully consider the data available and whether it is fit for purpose, model assumptions, the aforementioned considerations stated above regarding

population heterogeneity, differing dose response curves due to tumour response, dose levels, and chemicals, and the inherent stochastic variation in order to better account for the uncertainty.

Using a MOE approach to evaluate human risks from genotoxic carcinogens is based on the restricted observable dose-response data range from the (animal) experimental data. The risk to the MOE approach is that it is still dependent on the choice of model fit to the data. Even if the MOE limits itself to the observable range, the data is based on as few as three points in the middle range; it may therefore be misleading to express cancer risks in terms of predicted incidence.

This review emphasises the complex nature in understanding, study design and modelling of the multifaceted dose-response relationship for chemical carcinogenesis.

V.   Conclusions

The aim of this report was to evaluate evidence concerning the empirical form of the dose response relationship at low levels of exposure to genotoxic carcinogens in animals. The review of the relevant publications identified has highlighted some of the difficulties in determining the nature of this relationship.

Presentation of dose-response data

Visual inspection of dose-response data is often used as the first step in analysing the shape of the dose-response curve. However, it is readily apparent from the studies reviewed here that this in itself is a contentious issue, particularly in the so-called mega studies. This is because of the marked compression of data that will occur when a wide range of doses is plotted. Several methods of plotting dose-response data have been utilised, ranging from a linear-linear plot to a plot of dose on the Rozman scale versus linear response. Whilst there are critics of each of these methods, there does not appear to be a sound scientific rationale for choosing one over the other.

Dose-response model

Extensive dose-response modelling has been undertaken for some of the larger datasets described. In general, the models have been statistical rather than biologically motivated. The main difficulty in interpretation is that several models can give good fits to the data, but at best, as these in general span a tumour incidence range from 1-100%, they are not informative of the nature of the dose-response relationship at human relevant levels of exposure. The only exception to this is the large study undertaken in trout, where the incidence ranged from as low as 0.01%. This data set has not been subjected to such extensive analyses, as it was published only relatively recently. The results certainly show that the empirical form of the dose response curve is such that truncating it at higher doses/incidences such as those associated with conventional cancer bioassays in rodents would over-estimate the cancer risk at more human relevant exposures, due to the non-linear nature of the curve at lower doses.

Endpoints used to describe the dose-response relationship.

Whilst the key focus of this review was on cancer as an endpoint, several informative studies on pre-neoplastic effects were identified. These provided some mechanistic support for non-linearity of the tumour response, in that there was clear evidence of non-linearity in the dose-response relationship for such precursor effects, particularly in the appearance of

55

phenotypically altered cells. However, most of these studies were on hepatic effects, and the extent to which they are applicable to other sites is not clear. A number of studies have demonstrated that formation of DNA adducts can be linear to low doses, but that there is rarely a direct concordance between the level of adduction and tumour incidence. However, such studies are compounded by the differences in treatment time and cumulative dose necessarily used to investigate the different endpoints.

Implications for interpretation of the level of concern of the MOE

What is clear from the review of the various mega studies is that it has not been feasible to conduct a study, particularly in mammals, with sufficient power to adequately characterise the nature of the dose-response curve for a genotoxic carcinogen at the levels of exposure of most relevance to humans. Owing to current efforts to move away from the large scale use of animals in toxicity studies (e.g. The 3Rs: replacement, refinement and reduction; an ethical framework that promotes the humane use of animals [69]) it is highly unlikely that such mega studies will ever be conducted.

Efforts to continually re-analyse data produced from the mega studies conducted to date have only re-affirmed the inconsistencies and uncertainties inherent in visualising the data and in the statistical modelling techniques used in low dose response analysis. This was best highlighted by Purchase and Auton [2] in their reanalysis of the ED01 data or even by the approach proposed by Waddell [34, 35] in his reanalysis of the data; which all resulted in very different conclusions.

When one considers the inconsistencies observed across all the studies reviewed here, it is clear that there is a need to move away from the view that a single study (like the ED01 study) or series of studies (large scale or not) will provide definitive answers with regard to the shape of the dose response curve for all genotoxic carcinogens.

The publications reviewed here that sought to investigate the dose response analysis of known genotoxic carcinogens on preneoplastic effects [28, 31-33, 36, 37] demonstrate the growing efforts to derive conclusions about the low dose effects of such compounds on a case by case basis, with a more mechanistic foundation.

Threshold-seeking studies, such as that carried out by Gocke and Muller [37] in their evaluation of EMS and ENU, provide clear examples that thresholds do in fact appear to exist for some genotoxic carcinogens, based on detailed analyses of precursor effects. The effort therefore needs to be placed not only on the case by case analysis of genotoxic carcinogens but also on targeted chemical specific dose response analysis.

A well designed threshold seeking study that includes a comprehensive dose response analysis on a preneoplastic endpoint for a particular chemical would provide an additional dimension for the potential application of the MOE approach for characterising the risks from genotoxic carcinogens. The characterisation of carcinogenic risks based on evidence obtained

from the dose response analysis of a preneoplastic endpoint was attempted by Bailey et al [38] (albeit for illustrative purposes) who calculated a VSD based on DNA adduct data. The approach taken by Bailey et al [38] was to predict the tumour incidence response using the DNA adduct response data to predict the lower dose region of the tumour incidence response curve (<10.1 ppm) that corresponded to ED_{10}^{-6} in order to estimate the VSD.

Although Bailey et al [38] reached the conclusion that the DNA-adduct response data could not accurately predict the tumour dose response line, this conclusion was reached on the basis of the gradient differences observed for the two dose response lines (a slope of 1.31 was obtained for the DNA-adduct response line in contrast to the slope of 2.28 for the tumour incidence response line) rather than on the basis of any biological considerations.

When contemplating the use of such an approach to derive MOE values it is worth considering whether or not such an approach is scientifically justifiable on the basis of biological considerations. Taking the example of DNA adduct, it is not surprising that that the DNA-adduct response line was shown to be distinctly different from the tumour-incidence response line when one considers that the adduct response line is merely a snapshot of the overall carcinogenic process and does not take into account the full range of variables (biological processes) such as DNA repair that may occur further down the line or that the formation of DNA adducts may not be linked to the target gene potentially responsible for the formation of tumour cells at later time points in the genotoxic process.

The use of the preneoplastic-dose response data may not be appropriate for use in the prediction of the overall tumour incidence response but perhaps such data could be useful in the derivation of MOE values that are linked to a particular preneoplastic endpoint ("MOE-PNE": MOE based on a **pren**eoplastic **e**ndpoint). As such, a derived MOE value based on a preneoplastic endpoint would potentially correspond to a level of chemical exposure below which some of the key events that would normally lead to a genotoxic effect (characterised by tumour formation) would be unlikely to occur. A MOE- PNE would clearly have various caveats attached to the derived value to highlight the various uncertainties associated with any levels of concern associated and interpreted with such a value. It is unlikely that a value of >10,000 would be appropriate as a level of concern for such an endpoint.

A fundamental aspect to the potential derivation of a MOE-PNE would be a better understanding of the mode of action (MOA) for the genotoxic action for a particular chemical of interest. Inherent in this process would be the identification of some of the key events that must occur prior to the formation of tumours. Perhaps when we have a better understanding of the level of concern associated with a MOE derived on the basis of a tumour incidence-dose response as well as a better understanding of the various key stages of the genotoxic process, then we can be better equipped to pin point the exact time points for which a preneoplastic endpoint (e.g. DNA adduct or other preneoplastic endpoint) would be relevant for MOE derivation.

Targeted Chemical Specific Dose Response Analysis

As empirical dose response studies have insufficient power to determine the nature of the curve at low exposure levels, a more mechanistic approach is required. The endpoint investigated should be selected upon reflection and understanding of the mode of action of the chemical of interest, following an approach such as that suggested in the Key Events Dose-Response Framework (KEDRF) of the International Life Sciences Institute (ILSI) [70].

Sonich-Mullin et al. [71] defines "key events" as *measurable events that are critical to the induction of tumours as hypothesized in the postulated mode of action.* The identification of different thresholds for the effects of a genotoxic carcinogen for different preneoplastic endpoints is best illustrated by the work of Williams et al., [32] in their investigation of 2-AAF. This study provided a good example of the types of data that could be used in a key events dose response analysis (KEDRA). The pre-neoplastic endpoints studied ranged from DNA adduct formation, micronucleus induction/ chromosome damage, development of hepatic foci and levels of glutamine synthetase (GS+). That said, there needs to be greater emphasis on the characterisation of the dose response for intermediate endpoints that are the determining factors in the overall tumourigenic outcome for a chemical of interest. The KEDRF refers to such endpoints as "control points" in that they "engage specific mechanisms that may influence the ultimate outcome" by way of the magnitude or the probability of a tumourigenic outcome being observed [70]. Ultimately, a systems-based approach may provide the most rewarding means of addressing this complex issue.

Next steps:

1. Consideration of the implications for the derivation of MOEs based on preneoplastic endpoints.

    a. Selection of appropriate key preneoplastic endpoints associated with a tumour endpoint of interest.

    b. Consideration of the uncertainties that would be associated with such a value and how such information would be presented and interpreted.

2. Investigate the impact the shape of the dose response curve has on the derived MOE value and its associated level of concern.

References

1.	Littlefield, N.A., Farmer, J. H., Gaylor, D. W., and Sheldon, W. G, *Effects of dose and time in a long-term, low-dose carcinogenic study.* J Environ Pathol Toxicol, 1980. **3**(3 Spec No): p. 17-34.
2.	Purchase, I.F.H. and T.R. Auton, *Thresholds in Chemical Carcinogenesis.* Regulatory Toxicology and Pharmacology, 1995. **22**(3): p. 199-205.
3.	Jenkins, G.J.S., et al., *Genotoxic thresholds, DNA repair, and susceptibility in human populations.* Toxicology, 2009. **In Press, Corrected Proof**.
4.	Pottenger, L.H. and B.B. Gollapudi, *A case for a new paradigm in genetic toxicology testing.* Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 2009. **678**(2): p. 148-151.
5.	Zito, R., *Low doses and thresholds in genotoxicity: from theories to experiments.* J Exp Clin Cancer Res, 2001. **20**(3): p. 315-25.
6.	Weisburger, J.H., *Does the Delaney Clause of the U.S. Food and Drug Laws Prevent Human Cancers?* Fundamental and Applied Toxicology, 1994. **22**(4): p. 483-493.
7.	Greenstein, J., *Biochemistry of cancer.* Journal of the American Pharmaceutical Association, 1954. **43**(12): p. 769-770.
8.	Waddell, W.J., *Critique of dose response in carcinogenesis.* Human & Experimental Toxicology, 2006. **25**(7): p. 413-436.
9.	Ashby, J., *Change the rules for food additives.* Nature, 1994. **368**(6472): p. 582-582.
10.	Barlow, S., et al., *Risk assessment of substances that are both genotoxic and carcinogenic report of an International Conference organized by EFSA and WHO with support of ILSI Europe.* Food Chem Toxicol, 2006. **44**(10): p. 1636-50.
11.	Mantel, N.a.B., W.R., , *"Safety" testing of carcinogenic agents.* J. Nat. Cancer Inst. , 1961(27): p. pp. 455-470.
12.	Armitage, P., *The Assessment of Low-Dose Carcinogenicity.* Biometrics, 1982. **38**(ArticleType: research-article / Issue Title: Proceedings of "Current Topics in Biostatistics and Epidemiology." A Memorial Symposium in Honor of Jerome Cornfield / Full publication date: Mar., 1982 / Copyright © 1982 International Biometric Society): p. 119-129.
13.	Mantel, N., et al., *An Improved Mantel-Bryan Procedure for "Safety" Testing of Carcinogens.* Cancer Research, 1975. **35**(4): p. 865-872.
14.	Armitage, P.C., Theodore ed. *Dose Response Models in Risk Analysis*. Encyclopedia of Biostatistics. Vol. 6. 1998, Wiley: Chichester. 1219 - 1222.
15.	Armitage, P.C., Theodore ed. *Extrapolation, Low-Dose*. 1998, Wiley: Chichester. 1457-1461.
16.	Edler, L. and A. Kopp-Schneider, *Statistical models for low dose exposure.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 1998. **405**(2): p. 227-236.
17.	Gaylor, D.K., RL *Linear interpolation algorithm for low dose risk assessment of toxic substances.* ournal of Environmental Pathology and Toxicology 1980, (4): p. 305 - 312.
18.	Farmer JH, K.R.G.D., *Estimating and extrapolating of tumour probabilities from a mouse bioassay with survival/sacrifice components.* Risk Analysis, 1982(16): p. 395-398.
19.	Hogan, M.D., *Extrapolation of animal carcinogenicity data: limitations and pitfalls.* Environ Health Perspect., 1983. **47**: p. 333-7.
20.	(USEPA), E., *Appendix F: Lay Description of the Linearized Multistage Model*. 2012, USEPA: Washington DC.
21.	WHO, *Joint FAO/WHO Expert Committee on food Additivies. Sixty-fourth meeting, Rome, 8-17 February, 2005*. 2005.
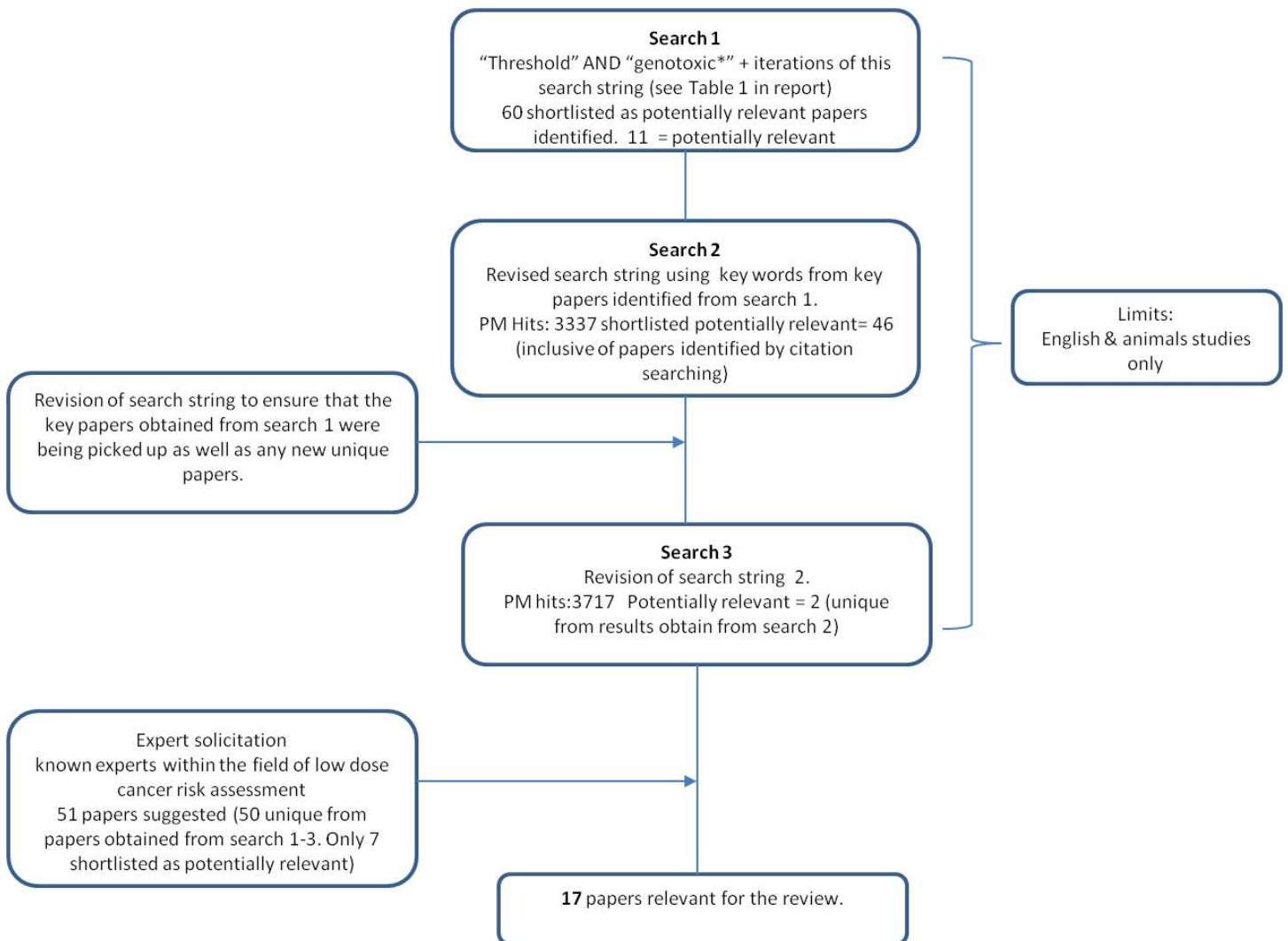
22. EFSA, *(European Food Safety Authority) Opinion of the Scientific Committee on a request from EFSA related to a Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic (Request No EFSA-Q-2004-020).* The EFSA Journal, 2005(282): p. 1-31.

23. Gaylor, D., Greenman DL, Frith CH, *Urinary bladder neoplasms induced in BALB/c female mice with low doses of 2-acetylaminofluorene.* J Environ Pathol Toxicol Oncol., 1985. **6**(1): p. 127-36.

24. SOT, *Re-examination of the ED01 Study: Review of Statistics: The Need for Realistic Statistical Models for Risk Assessment* Toxicological Sciences, 1981. **1**(1): p. 124-126

25. Zapponi GA, B.A., Lupi C., *Reproducibility of low-dose extrapolation procedure: comparison of estimates obtained using different rodent species and strains.* Biomed Environ Sci., 1988. **1**(2): p. 160-70.

26. Peto, R., et al., *Effects on 4080 Rats of Chronic Ingestion of N-Nitrosodiethylamine or N-Nitrosodimethylamine: A Detailed Dose-Response Study.* Cancer Research, 1991. **51**(23 Part 2): p. 6415-6451.

27. Poirier, M.C., et al., *Comparison between DNA adduct formation and turmorigenesis in livers and bladders of mice chronically fed 2-acetylaminofluorence.* Carcinogenesis, 1991. **12**(5): p. 895-900.

28. Williams, G.M., et al., *Non-linearity of neoplastic conversion induced in rat liver by low exposures to diethylnitrosamine.* Carcinogenesis, 1993. **14**(10): p. 2149-2156.

29. Choy, W., *A review of the dose-response induction of DNA adducts by aflatoxin B1 and its implications to quantitative cancer-risk assessment.* Mutat Res., 1993. **296**(3): p. 181-98.

30. Shirai T, T.S., Sano M, Masui T, Hasegawa R, Ito N., *Carcinogenicity of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) in rats: dose-response studies.* Princess Takamatsu Symp, 1995. **23**: p. 232-9.

31. Williams, G., Iatropoulos, MJ., Jeffrey, AM., Luo F, Wang, C., Pittman B., *Diethylnitrosamine exposure-responses for DNA ethylation, hepatocellular proliferation, and initiation of carcinogenesis in rat liver display non-linearities and thresholds.* Arch Toxicol., 1999. **73**(7): p. 394-402.

32. Williams, G.M., M.J. Iatropoulos, and A.M. Jeffrey, *Thresholds for the Effects of 2-Acetylaminofluorene in Rat Liver.* Toxicologic Pathology, 2004. **32**(2 suppl): p. 85-91.

33. Williams, G.M., M.J. Iatropoulos, and A.M. Jeffrey, *Mechanistic Basis for Nonlinearities and Thresholds in Rat Liver Carcinogenesis by the DNA-Reactive Carcinogens 2-Acetylaminofluorene and Diethylnitrosamine.* Toxicologic Pathology, 2000. **28**(3): p. 388-395.

34. Waddell, W.J., *Thresholds of Carcinogenicity in the ED01 Study.* Toxicological Sciences, 2003. **72**(1): p. 158-163.

35. Waddell, W.J., *Threshold for carcinogenicity of N-nitrosodiethylamine for esophageal tumors in rats.* Food and Chemical Toxicology, 2003. **41**(5): p. 739-741.

36. Doak, S.H., et al., *Mechanistic Influences for Mutation Induction Curves after Exposure to DNA-Reactive Carcinogens.* Cancer Research, 2007. **67**(8): p. 3904-3911.

37. Gocke, E. and L. Müller, *In vivo studies in the mouse to define a threshold for the genotoxicity of EMS and ENU.* Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 2009. **678**(2): p. 101-107.

38. Bailey, G.S., et al., *Nonlinear Cancer Response at Ultralow Dose: A 40800-Animal ED001 Tumor and Biomarker Study.* Chemical Research in Toxicology, 2009. **22**(7): p. 1264-1276.

39. Gaylor, D.W., *The ED01 study: summary and conclusions.* J Environ Pathol Toxicol, 1980. **3**(3 Spec No): p. 179-83.

40. Cairns, T., *The EDO1 study: Introduction, objectives and Experimental design.* Journal of Environmental Pahtology and Toxicology, 1979(3): p. 1-7.

41. Mehlma, J.A.S.M.A., *Innovations in cancer risk assessment (ED01 study).* Food and Cosmetics Toxicology, 1981. **19**: p. 776-776.

42. DJ, F., *Probit Analysis*.

. 1971: Cambridge University Press.

43. Farmer JH, K.R., Greenman DL, *Dose and Time Response Models for the Incidence of Bladder and Liver Neoplasms in Mice fed 2-Acetylaminofluorene Continuously.* Journal of Environmental Pahtology and Toxicology, 1979(3): p. 55-68.

44. Farmer, J.H., Kodell, R. L., Greenman, D. L., and Shaw, G. W, *Dose and time responses models for the incidence of bladder and liver neoplasms in mice fed 2-acetylaminofluorene continuously.* J Environ Pathol Toxicol, 1980. **3**(3 Spec No): p. 55-68.

45. SOT, *Re-examination of the ED01 Study: Why the Society of Toxicology Became Involved* Toxicological Sciences, 1981. **1**(1): p. 27.

46. SOT, *Re-examination of the ED01 Study: Summary and Conclusions* Toxicological Sciences, 1981. **1**(1): p. 127-128.

47. SOT, *Members of the Society of Toxicology: ED01 Task Force.* Toxicological Sciences, 1981. **1**(1): p. 26.

48. Waddell, W.J., *Thresholds of Carcinogenicity of Flavors.* Toxicological Sciences, 2002. **68**(2): p. 275-279.

49. Rozman, K.K., *Quantitative definition of toxicity: a mathematical description of life and death with dose and time as variables.* Medical Hypotheses, 1998. **51**(2): p. 175-178.

50. Crump, K.S. and H.J. Clewell, *Letter.* Toxicological Sciences, 2003. **74**(2): p. 485.

51. Andersen, M.E., R.B. Conolly, and D.W. Gaylor, *Letter.* Toxicological Sciences, 2003. **74**(2): p. 486-487.

52. Lutz, W.K., *Letter.* Toxicological Sciences, 2003. **75**(1): p. 223.

53. Haseman, J.K., *An Alternative Perspective: A Critical Evaluation of the Waddell Threshold Extrapolation Model in Chemical Carcinogenesis.* Toxicologic Pathology, 2003. **31**(5): p. 468-470.

54. Haseman, J.K., *Response to Waddell & Rozman.* Toxicologic Pathology, 2003(31): p. 715-716.

55. Waddell, *Letters to the Editor.* Tox Sci 2003. **74**: p. 485-488.

56. Peto, R., et al., *Dose and Time Relationships for Tumor Induction in the Liver and Esophagus of 4080 Inbred Rats by Chronic Ingestion of N-Nitrosodiethylamine or N-Nitrosodimethylamine.* Cancer Research, 1991. **51**(23 Part 2): p. 6452-6469.

57. Appleton BS, G.M., Campbell TC., *Linear dose-response curve for the hepatic macromolecular binding of aflatoxin B1 in rats at very low exposures.* Cancer Res., 1982. **42**(9): p. 3659-62.

58. Gebhardt, R., T. Tanaka, and G.M. Williams, *Glutamine synthetase heterogeneous expression as a marker for the cellular lineage of preneoplastic and neoplastic liver populations.* Carcinogenesis, 1989. **10**(10): p. 1917-1923.

59. Williams, G.M., *The pathogenesis of rat liver cancer caused by chemical carcinogens.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1980. **605**(2): p. 167-189.

60. Lutz, W.K. and R.W. Lutz, *Statistical model to estimate a threshold dose and its confidence limits for the analysis of sublinear dose-response relationships, exemplified for mutagenicity data.* Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 2009. **678**(2): p. 118-122.

61. Rozman, K.K., *Rebuttal to Haseman.* Toxicologic Pathology, 2003. **31**(6): p. 714.

62. Kodell RL, G.D., Greenman DL, Littlefield NA, Farmer JH, *Response to the society of toxicology task force re-examination of the ED01 study.* Fundam Appl Toxicol., 1983. **3**(4): p. 3a-12a.

63. Rozman, K.K., et al., *A Toxicologist's View of Cancer Risk Assessment.* Drug Metabolism Reviews, 1996. **28**(1-2): p. 29-52.

64. Waddell WJ., F.S., Williams GM, *Concordance of thresholds for carcinogenicity of N-nitrosodiethylamine.* Arch Toxicol., 2006. **80**(6): p. 305-9. Epub 2005 Nov 25.

65. Fukushima, S., et al., *Lack of a dose-response relationship for carcinogenicity in the rat liver with low doses of 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline or N-nitrosodiethylamine.* Japanese journal of cancer research : Gann, 2002. **93**(10): p. 1076-1082.

66. Rozman, K., D. John, and R. William, *Influences of Dynamics, Kinetics, and Exposure on Toxicity in the Lung*, in *Toxicology of the Lung, Fourth Edition*. 2005, Informa Healthcare. p. 195-229.

67. Waddell, W.J., Fukushima S., Williams GM., *Concordance of thresholds for carcinogenicity of N-nitrosodiethylamine.* Arch Toxicol., 2006. **80**(6): p. 305-9. Epub 2005 Nov 25.

68. Waddell, W.J., *Analysis of thresholds for carcinogenicity.* Toxicology Letters, 2004. **149**(1-3): p. 415-419.

69. Robinson, V., *Finding alternatives: an overview of the 3Rs and the use of animals in research.* School Science Review, 2005. **87**: p. 319.

70. Julien, E., et al., *The Key Events Dose-Response Framework: A Cross-Disciplinary Mode-of-Action Based Approach to Examining Dose-Response and Thresholds.* Critical Reviews in Food Science and Nutrition, 2009. **49**(8): p. 682-689.

71. Sonich-Mullin, C., et al., *IPCS Conceptual Framework for Evaluating a Mode of Action for Chemical Carcinogenesis.* Regulatory Toxicology and Pharmacology, 2001. **34**(2): p. 146-152.

Appendix I

Flow diagram to show the search strategy employed to identify relevant papers for the review

Search 1
"Threshold" AND "genotoxic*" + iterations of this search string (see Table 1 in report)
60 shortlisted as potentially relevant papers identified. 11 = potentially relevant

Search 2
Revised search string using key words from key papers identified from search 1.
PM Hits: 3337 shortlisted potentially relevant= 46 (inclusive of papers identified by citation searching)

Revision of search string to ensure that the key papers obtained from search 1 were being picked up as well as any new unique papers.

Limits:
English & animals studies only

Search 3
Revision of search string 2.
PM hits:3717 Potentially relevant = 2 (unique from results obtain from search 2)

Expert solicitation
known experts within the field of low dose cancer risk assessment
51 papers suggested (50 unique from papers obtained from search 1-3. Only 7 shortlisted as potentially relevant)

17 papers relevant for the review.

Data collection and study design methods of three mega studies

Information regarding data collection and study design methods of three studies have been extracted to determine the data structure and effectiveness of the studies to be used for (a) extrapolation from high to low dose-response, and (b) analysis for human cancer risk assessments using MoE.

Table 6: Study design methods for the three mega studies reviewed

		NCTR [1]	Peto [26]	Bailey [38]
Purpose				
Animal	*Type*	Mice	Rats	Rainbow Trout
	No	24, 192	4080	40,800
	Per group	4/cage	10/group	100/tank
	Population	Female only	240×2 Males and 240×2 Females	Presumably males and females; not stated explicitly
	Other	3-4 weeks starting age	6 weeks, inbred Colworths	*O. mykiss* Shasta strain, possibly 18 weeks at start – 1.5kg (from previous paper), but not stated
Chemical	*Name*	2-AAF	N-nitrosodiethylamine & N-nitrosodimethylamine	Dibenzo[*a,l*]pyrene
	Dose Levels	0, 30, 35, 45, 60, 75, 100, & 150 ppm	5 levels between 0 and 1	0, 0.45, 1.27, 3.57, 10.1, 28.4, 80 and 225 ppm
	No. Treatments	80 + 1 control	15 + 1 control, 6 replicates	8×4 replicates with 10 000 controls, 102 replicates
	Dose rate/process	Ingested through food?	Water	Incorporated into diet (Oregon Test Diet). Treated for 4 weeks.

	Other		60 rats per level	100 fish per replicate
Response	Type	Liver and bladder neoplasms	Esophagus/liver & lung/nasophargne	Liver and stomach neoplasms
	Baseline rate	Control, excess of 1%	Control	Liver: 0.00085-0.00195; stomach: 0.00043-0.00236
	Measurement	Post mortem	Post mortem	Post mortem
	Measurement interval	9, 12, 13, 14, 15, 18, 24, 33 months	$1/10^{th}$ @ 12 mo, $1/10^{th}$ @ 18mo,	50 fish each on days 15 and 29 sampled for DNA addicts
	Measurement Time Scale	All sacrificed at 33 months	Till death	Terminated at 9 months